Investigation of The Genetic Regulation of Delayed Puberty

Dr Sasha Howard

MA(Cantab), MBBS, MRCPCH, MSc

Submitted in partial fulfilment of the requirements of the Degree of Doctor of Philosophy

Centre for Endocrinology

William Harvey Research Institute

Barts and the London School of Medicine and Dentistry

Queen Mary University of London

Statement of Originality

I, Sasha Howard, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature: 

Date: 20/01/2017

Publications arising from this thesis

Research Manuscripts:

Howard SR, Guasti L, Ruiz-Babot G, Mancini A, David A, Storr HL, Metherell LA, Sternberg MJE, Cabrera CP, Warren HR, Barnes MR, Quinton R, de Roux N, Young J, Guiochon-Mantel A, Wehkalampi K, André V, Gothilf Y, Cariboni A, Dunkel L (2016) *IGSF10* mutations dysregulate gonadotropin-releasing hormone neuronal migration resulting in delayed puberty. EMBO Mol Med, 8 (6): 626-642, doi: 10.15252/emmm.201606250 (Appendix 8)

- Selected as cover article for the same issue (Appendix 7)

Joustra S, Wehkalampi K, Oostdijk W, Biermasz N, **Howard S**, Silander T, Bernard D, Wit J, Dunkel L, Losekoot M (2015) *IGSF1* variants in boys with familial delayed puberty. Eur J Pediatr, 174 (5): 687-692. doi: 10.1007/s00431-014-2445-9

Poliandri A, Miller D, **Howard S**, Nobles M, Harmer S, Tinker A, McKay T, Guasti L, Dunkel L (2017). Generation of kisspeptin-responsive GnRH neurons from human pluripotent stem cells. Mol Cell Endocrinol, Feb 21. pii: S0303-7207(17)30117-X. doi: 10.1016/j.mce.2017.02.030. [Epub ahead of print]

Reviews/ Book Chapters:

Howard S and Dunkel L (2014) Delayed puberty and hypogonadism, Female. “Reference Module in Biomedical Sciences”, Elsevier. 10.1016/B978-0-12-801238-3.03882-4.

Howard S and Dunkel L (2016) Sex steroid and gonadotropin treatment in male delayed puberty. In: "Puberty from Bench to Clinic; Lessons for Clinical Management of Pubertal Disorders" (ed. by J.P. Bourguignon and A.S. Parent, Karger), Endocr Dev, 29:185-97. doi: 10.1159/000438891

Huhtaniemi IT, **Howard S**, Dunkel L & Anderson RA (2017) The Gonadal Axis: A Life Perspective. In: Pfaff, D.W and Joëls, M. (editors-in-chief), Hormones, Brain, and Behavior 3rd edition, Vol 4. Oxford: Academic Press; pp. 3–58.

Howard S and Dunkel L (2017) Normal and Delayed Puberty. In: Winters, S. & Huhtaniemi, IT. (Springer) Male Hypogonadism: Basic, Clinical and Therapeutic Principles 2nd edition, in print.

Acknowledgements

I would like to thank: my supervisors Professor Leo Dunkel and Dr Helen Storr for their guidance, encouragement and mentorship. I have been privileged to undertake my PhD in an inspiring and enthusiastic environment that has allowed me to become experienced across a number of different endocrine and genetic disciplines and to develop many translational skills for my future career.

The patients and their families, without whom this work would not have been possible.

Dr Leonardo Guasti, Dr Ariel Poliandri and Alessandra Mancini, who as part of our puberty research group have taught, advised, assisted and appraised my work throughout my PhD.

Dr Lou Metherell, Dr Claudia Cabrera and Dr Michael Barnes for their bioinformatics assistance and instruction, without which this project would not have been possible.

Our collaborators: Dr Anna Cariboni and Dr Valentina Andre and their group for their collaborative work on both IGSF10 and HS6ST1; Dr Giles Yeo and Dr Antony Coll, Dr Ken Ong, Dr John Perry and their groups for their collaborative work on FTO; Dr Richard Quinton, Professor Nicolas de Roux, Professor Jacques Young and Dr Anne Guiochon-Mantel for providing clinical samples; Professor Yoav Gothilf for zebrafish work on IGSF10; Dr Alessia David and Professor Michael Sternberg for *in silico* analysis; and Dr Gerard Ruiz Babot and Dr Helen Warren for laboratory and statistical support respectively.

My funding bodies: this work was supported by a Clinical Training Fellowship from Barts and the London Charity (grant number 417/1551), a Clinical Research Training Fellowship from the Wellcome Trust (grant number 102745) and by the Rosetrees Trust (grant number M222).

Finally, I would like to thank my family and my husband, James, and children, Bella, Ted and Lottie, for their support, encouragement and distraction from the concerns of work.

Abstract

The genetic control of puberty remains an important but mostly unanswered question. Late pubertal timing affects over 2% of adolescents and is associated with adverse health outcomes. Self-limited delayed puberty (DP) segregates in an autosomal dominant pattern and is highly heritable; however, its neuroendocrine pathophysiology and genetic regulation remain unclear.

Our large, accurately phenotyped cohort of patients with familial self-limited DP is a unique resource with a relatively homogeneous genetic composition. I have utilised this cohort to investigate the genetic variants segregating with the DP trait in these pedigrees. Whole exome sequencing in eighteen probands and their relatives, and subsequent targeted sequencing in an extended subgroup of the cohort, has revealed potential novel genetic regulators of pubertal timing. In ten unrelated probands, I identified rare mutations in *IGSF10*, a gene that is strongly expressed in the nasal mesenchyme during embryonic migration of gonadotropin-releasing hormone (GnRH) neurons. *IGSF10* knockdown both *in vitro* and in a transgenic zebrafish model resulted in perturbed GnRH neuronal migration. Loss-of-function mutations in *IGSF10* were also identified in five patients with absent puberty due to hypogonadotropic hypogonadism (HH). Additionally, I have identified and investigated one rare, pathogenic mutation in *HS6ST1* – a gene known to cause HH - in one family with DP, and two rare variants in *FTO* – a gene implicated in the timing of menarche in the general population - in 3 families. Further potentially pathogenic variants have emerged from investigating candidate genes identified from microarray studies (*LGR4*, *SEMA6A* and *NEGR1*) and from related clinical phenotypes (*IGSF1*).

My findings serve to highlight the fascinating heterogeneity of genetic defects resulting in disordered puberty. Although our understanding of this highly complex biological network remains imperfect, my results demonstrate the importance of defects in GnRH neuronal development and function, and potentially body mass regulation, in the control of pubertal timing.

Table of Contents

List of Figures

List of Tables

List of abbreviations

µM	micromolar
A1-3	Tanner axillary hair stage 1-3
AAM	age at menarche
ARC	arcuate nucleus
AVPV	anteroventral periventricular nucleus
B1-5	Tanner breast stage 1-5
BMI	body mass index
CAH	congenital adrenal hyperplasia
CCP	central precocious puberty
CDGP	constitutional delay in growth and puberty
CNS	central nervous system
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dNTPs	deoxynucleotidetriphosphates
DP	delayed puberty
Dpf	days post fertilisation
E	embryonic day
E2	estradiol
E Coli	*Escherichia* coli
EDTA	Trypsin/Ethylenediaminetetraacetic acid
FSH	follicular stimulating hormone
G1-5	Tanner genital stage 1-5
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GFP	green fluorescent protein
GH	growth hormone
GnRH	gonadotropin releasing hormone
GWAS	genome wide association studies
HA	hypothalamic amenorrhoea
hCG	human chorionic gonadotropin
HH	hypogonadotropic hypogonadism
HLOD	heterogeneity logarithm of the odds
Hpf	hours post fertilisation
HPG	hypothalamic-pituitary-gonadal
HSDS	height standard deviation score
IGD	isolated gonadotropin deficiency
IGF	insulin-like growth factor
ISH	in situ hybridisation
Kb	kilobase pairs
KD	knockdown
kDa	kilodalton
KS	Kallmann syndrome
LCH	Langerhans cell histiocytosis
LH	luteinising hormone
M	menarche

MAS	McCune Albright syndrome
mL	millilitre
mM	millimolar
M-MLV	Maloney murine leukaemia virus
mRNA	messenger RNA
m.s.a.	multiple sequence alignment
NGS	next generation sequencing
OMIM	online mendelian inheritance in man
P	progesterone
P1-5	Tanner pubic hair stage 1-5
PBS	Phosphate-buffered saline
PCR	polymerase chain reaction
pcw	post-conceptual weeks
pg	picograms
PHV	peak height velocity
qPCR	quantitative PCR
RIPA	Radioimmunoprecipitation assay
RNA	Ribonucleic acid
RNase	Ribonuclease
Rpm	rotations per minute
RT	room temperature
RT-qPCR	real time qPCR
SD	standard deviation of the mean
SDS	standard deviation score
SDS-Page	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
shRNA	small hairpin RNA
siRNA	small interfering RNA
SNP	single nucleotide polymorphism
TH	target height
Tm	melting temperature
WES	whole exome sequencing
WT	wild-type
yrs	years

Chapter 1: Introduction

1.1 Puberty and the HPG axis

Puberty is the maturational process of the reproductive endocrine system that results in achievement of adult height and body proportion, in addition to development of the genital organs and the capacity to reproduce. The onset of puberty is driven by an increase in the pulsatile release of gonadotropin-releasing hormone (GnRH) from the hypothalamus. This activation results in increased luteinizing hormone (LH) and follicular-stimulating hormone (FSH) release from the anterior pituitary, which act on the gonads to stimulate their development, gametogenesis and sex steroid production (Fig 1.1).

The development of the hypothalamic-pituitary-gonadal (HPG) axis is exceptional in that GnRH neurons develop in metazoan embryos outside of the central nervous system. Immature GnRH precursor neurons are first detectable in the olfactory placode in the nose from an early embryological stage (E10.5-E11 in mice), and then begin a complex journey towards the hypothalamus (1).

The axis is active in fetal and in early infant life, the so-called 'mini-puberty', and then becomes dormant with levels of LH and FSH low or undetectable in young children between the age of 2 and 8-9 years (2). Development of the clinical features of puberty is initiated by the reactivation of the HPG axis after this relative quiescence during childhood. What drives this suppression of the axis during childhood, and what controls the release of this 'brake' and the timing at which this occurs, is little understood.

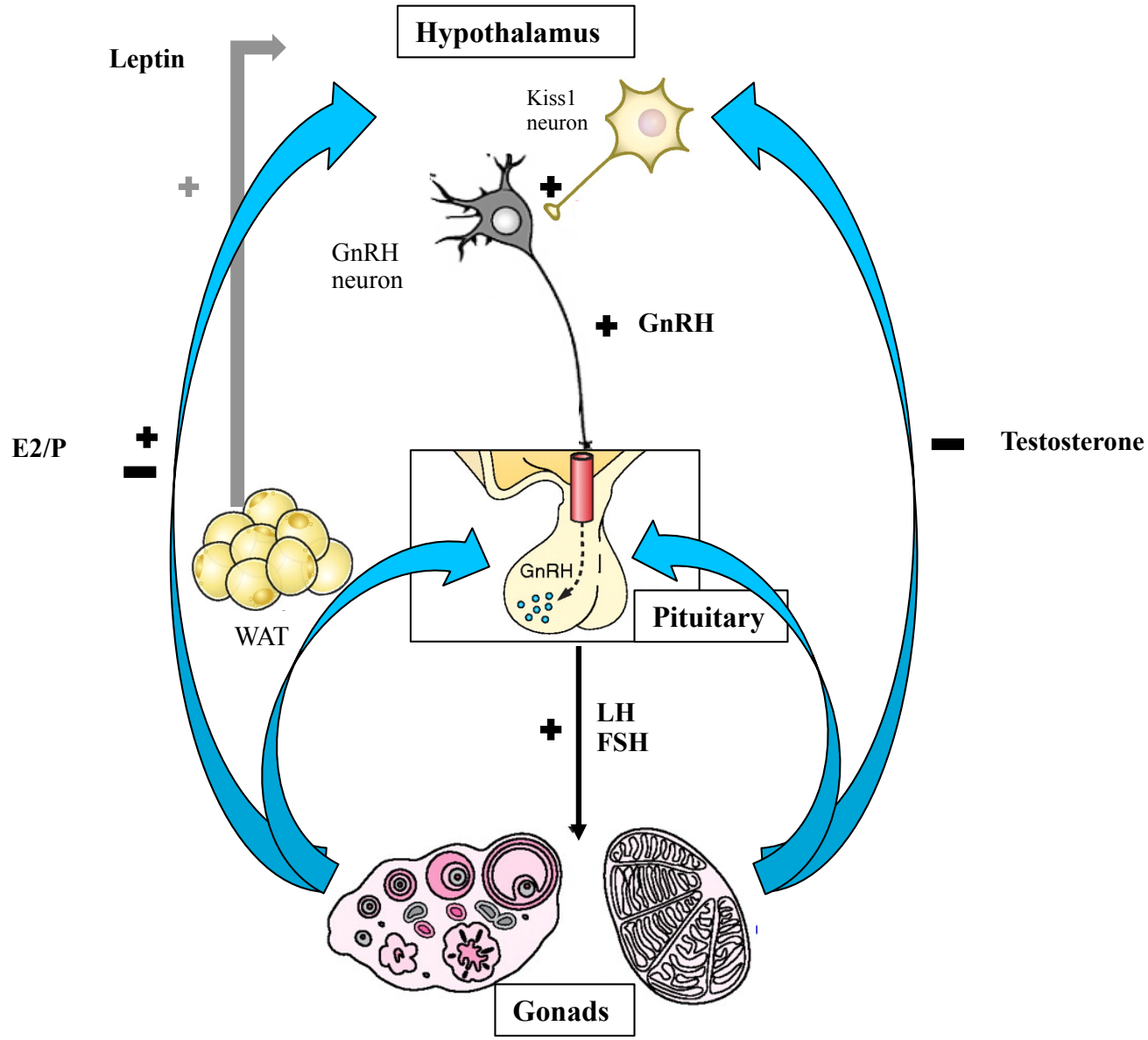


Figure 1.1 – The hypothalamic-pituitary-gonadal axis. WAT – white adipose tissue; E2/P – estradiol/progesterone

Once the adult HPG axis is fully functional it remains responsive to external influences, with variation in measured gonadotropins and sex steroids in response to nutrition, stress and other environmental factors, as well as variation in females through the menstrual cycle. Thus, the successful functioning of the HPG axis is both vital for reproductive health but also requires phenomenal plasticity during different stages of life and in differing circumstances.

1.2 Clinical Markers of Puberty

In boys the first physical marker of pubertal onset is an increase in testis volume above the pre-pubertal volume of 3 mL (Tanner stage G2) (3) (Fig. 1.2). In girls the first physical sign that true puberty has commenced is the onset of breast development (Tanner stage B2) (4) (Fig. 1.2).

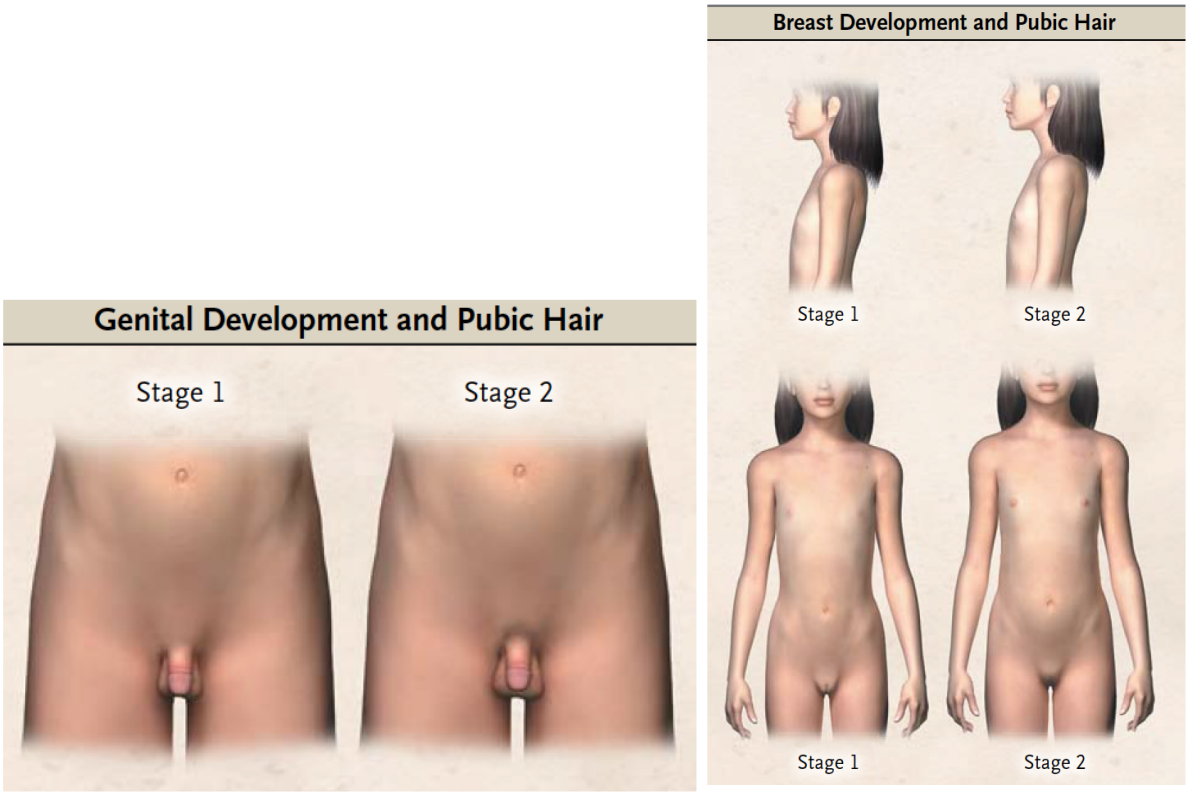


Figure 1.2 – Tanner staging of puberty onset in boys and girls (5) Reproduced with author permission.

In both genders, there is a classic pattern of pubertal progression, with breast and pubic hair development in girls, and in boys the onset of testicular enlargement followed by pubic hair development and penile growth (Table. 1.2). The tempo and process of puberty are well conserved across ethnicities and populations.

Tanner Stage	**Females**		**Males**	
	Breast	Pubic hair	Genitalia	Pubic hair
1	Nipple elevation only	None	Testis 1-3ml	None
2	Small raised breast bud	Growth along labia, sparse, lightly pigmented	Testis >3ml Scrotal enlargement	Sparse, lightly pigmented
3	Breast and areola enlarge with no contour difference	Increases in amounts, darkens, starts to curl	Testis continue to enlarge, penis lengthens	Increases in amounts, darkens, starts to curl
4	Further enlargement with nipple and areola projecting to form secondary mound	Resembles adult type but not spread to medial thighs	Scrotum darkens, widening of glans penis	Resembles adult type but not spread to medial thighs
5	Adult contour with areola and breast in same contour, nipple projecting	Spreads to medial thighs, adult distribution	Adult size and morphology	Spreads to medial thighs, adult distribution

Table 1.2 – Details of the Tanner stages of puberty

1.3 Timing of Puberty

In the general population, there is a near-normal distribution of the timing of puberty, with the mean age of onset of G2 at 11.5yrs in boys (Fig. 1.3.1) and 11yrs for B2 in girls (Fig. 1.3.2). Within this distribution there has been in recent years an increasing degree of skew at both ends of the spectrum, as an earlier age of pubertal onset (B2 or G2) has become more prevalent, as well as an increase in the number of children completing their puberty at a later age (6).

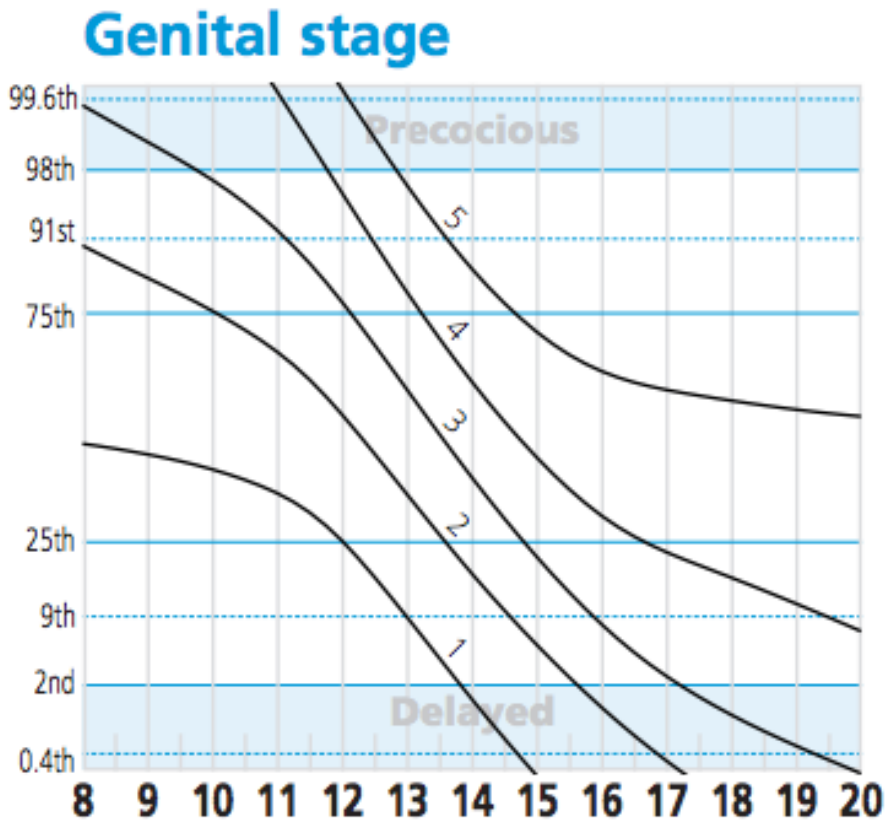


Figure 1.3.1 – The distribution of pubertal timing in healthy boys. These data have been incorporated into UK growth charts and are available at www.growthcharts.rcpch.ac.uk. Original data from (7)

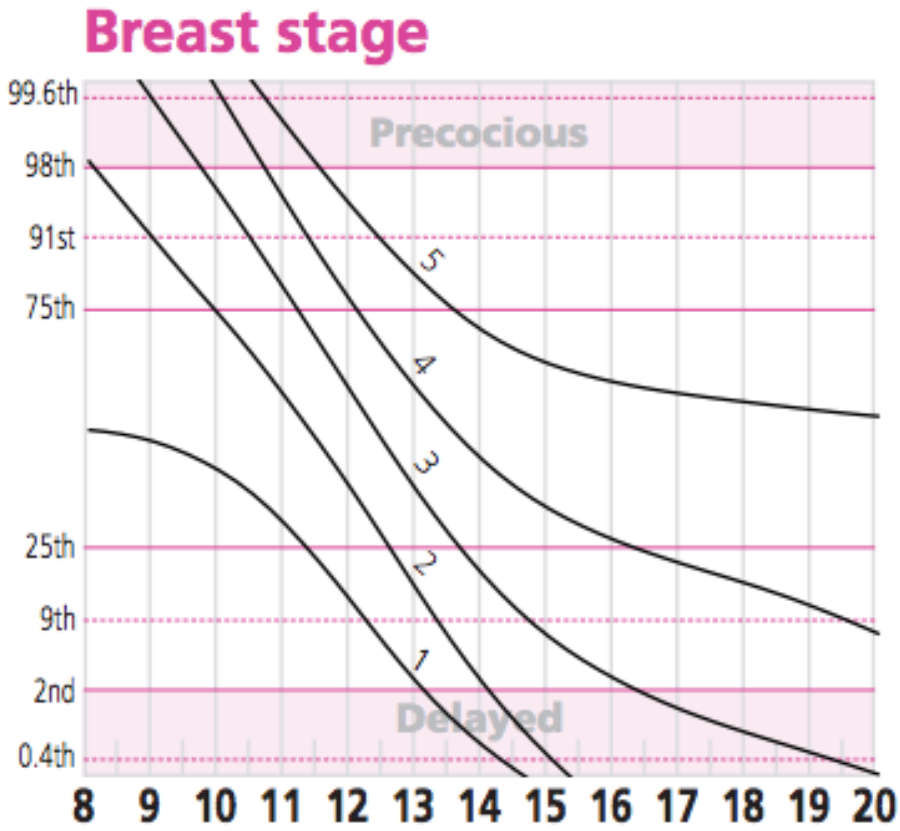


Figure 1.3.2 – The distribution of pubertal timing in healthy girls. These data have been incorporated into UK growth charts and are available at www.growthcharts.rcpch.ac.uk. Original data from (7)

In healthy boys the normal age limits for G2 development are between 9 and 14 years (5). The great majority of Caucasian girls have at least early signs of secondary sexual development by 13 years of age (Fig. 1.3.2). Whilst a large variability in the timing of pubertal onset exists in both genders, clear age cut-offs for normal pubertal development have been drawn. However, for both genders the age limits for identifying children who need evaluation for precocious or delayed puberty (DP) may vary in different ethnic groups.

Peak height velocity (PHV), and peak pubertal growth hormone production, coincides approximately with mid-point of pubertal development (Fig. 1.3.3 and Fig. 1.3.4). Up to 25% of total adult height is achieved from growth during puberty, but the amplitude and peak velocity of the pubertal growth spurt is not fixed and varies with age at onset of puberty (8).

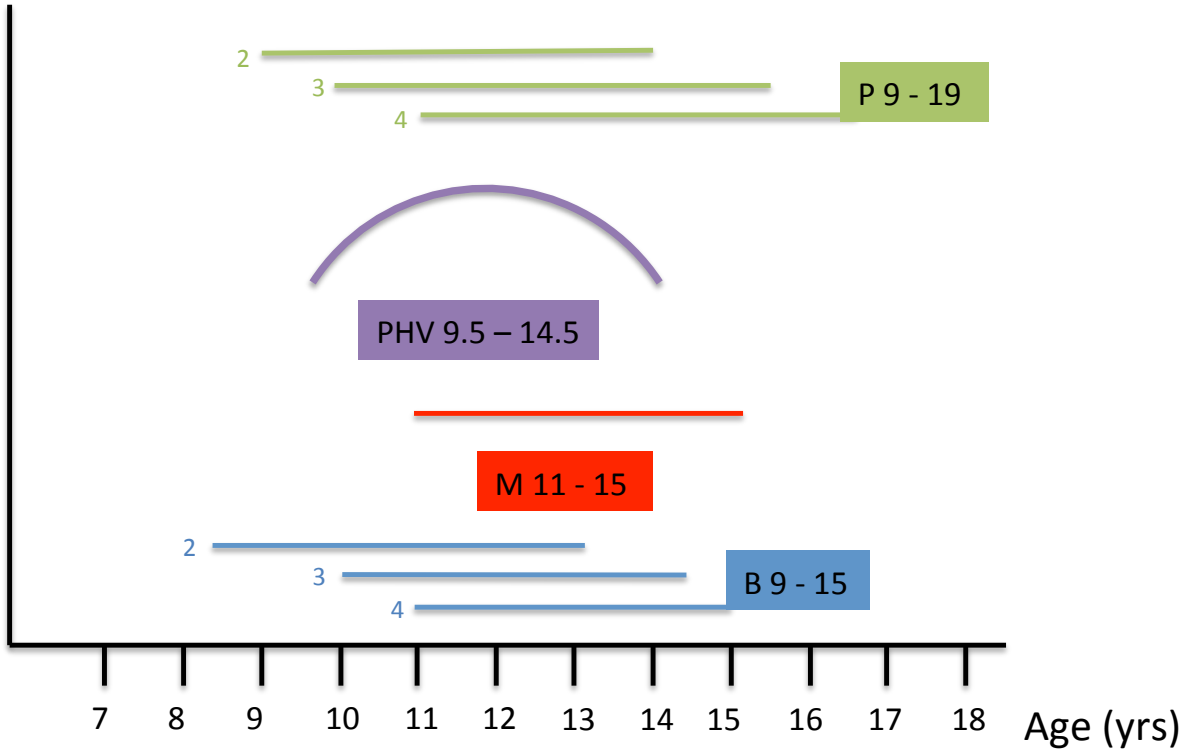


Figure 1.3.3 – Relationship between peak height velocity and pubertal development (female). P – pubic hair stage; PHV – peak height velocity; M – menarche; B – breast stage. Age ranges presented: in coloured boxes – 3rd-97th centiles; in linear format – 3rd-97th centiles for each tanner stage e.g. B2; curved lines represent pubertal growth spurt with age ranges for PHV. Original concept from(9). Data from (7, 10)

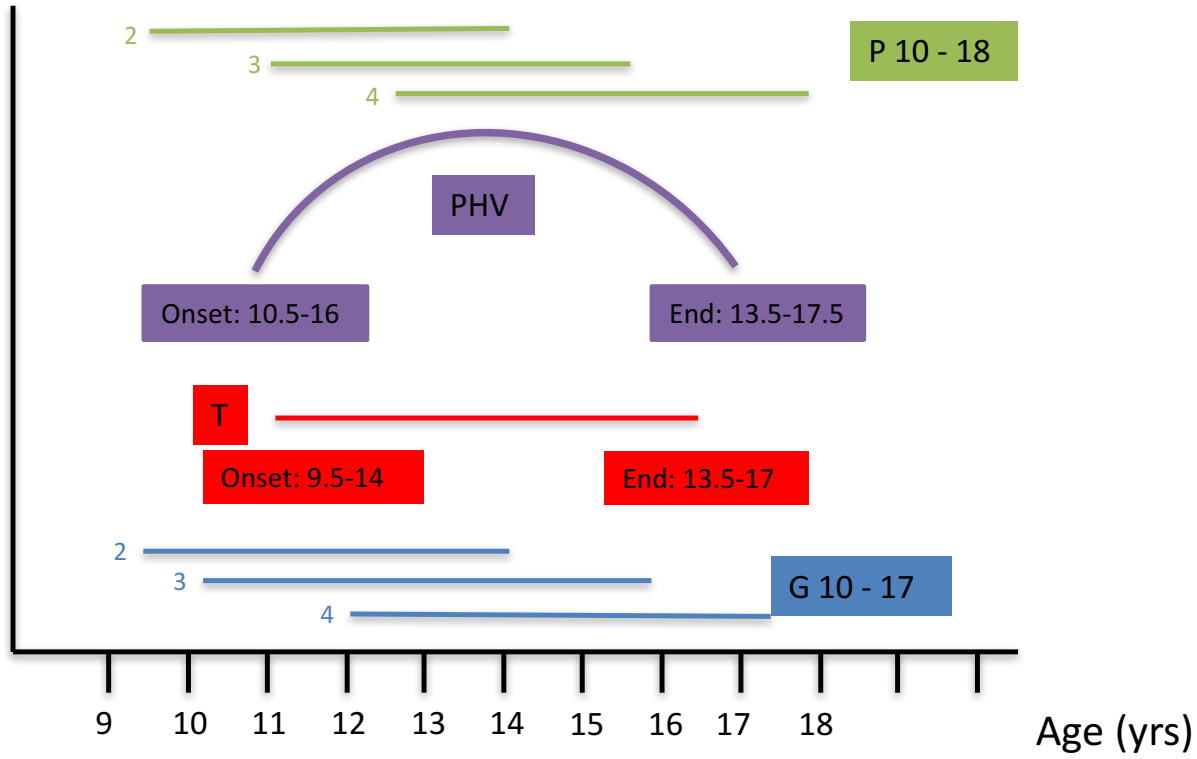


Figure 1.3.4 – Relationship between peak height velocity and pubertal development (male). P – pubic hair stage; PHV – peak height velocity; T – testicular growth; G – genital stage. Age ranges presented: in linear format – 3rd-97th centiles for each Tanner stage e.g. G2; curved lines represent pubertal growth spurt with age ranges for onset and completion; in coloured boxes – age range in years for onset to completion of each parameter. Original concept from (9). Data from (7, 10)

In addition, a clear gender discrepancy exists in both the age at and degree of PHV attained. In girls PHV occurs at B2-3, at an average age of 11.5 yrs with a PHV of 8.3cm/yr. In contrast, in boys the PHV coincides with G3-4 at an average age of 13.5yrs and achieves an incremental rate of 9.5cm/yr (11).

The effect of the timing of puberty onset on pubertal growth is measurable: early puberty is associated with a large pubertal growth spurt whilst late maturers who have a longer pre-pubertal period of growth in turn experience a less pronounced pubertal growth spurt (12). Therefore, there is regulatory compensation at work: although extremes of pubertal timing potentially lead to a degree of final height reduction due to the reduced overall period of growth in precocious puberty, or poor PHV in DP (13), in general within wide limits the timing of puberty does not greatly influence adult height.

1.4 Regulators of the Timing of Puberty

This variability in the timing of puberty in healthy adolescents is governed by complex regulatory mechanisms involving genetic, environmental and other factors (14). Nutritional status, adoption, geographical migration and emotional well-being have all been shown to have an effect on pubertal timing (15-17). The timing of puberty in the developed world in most countries exhibited a rapid decrease, most noted in girls, in the first half of the 20th century (18, 19) (Fig. 1.4.1). In boys, these trends have been less clearly shown. A small but significant change in the normal spectrum of timing of G2 development has been documented in a European cohort (20) and in the US (21), but remains controversial (22).

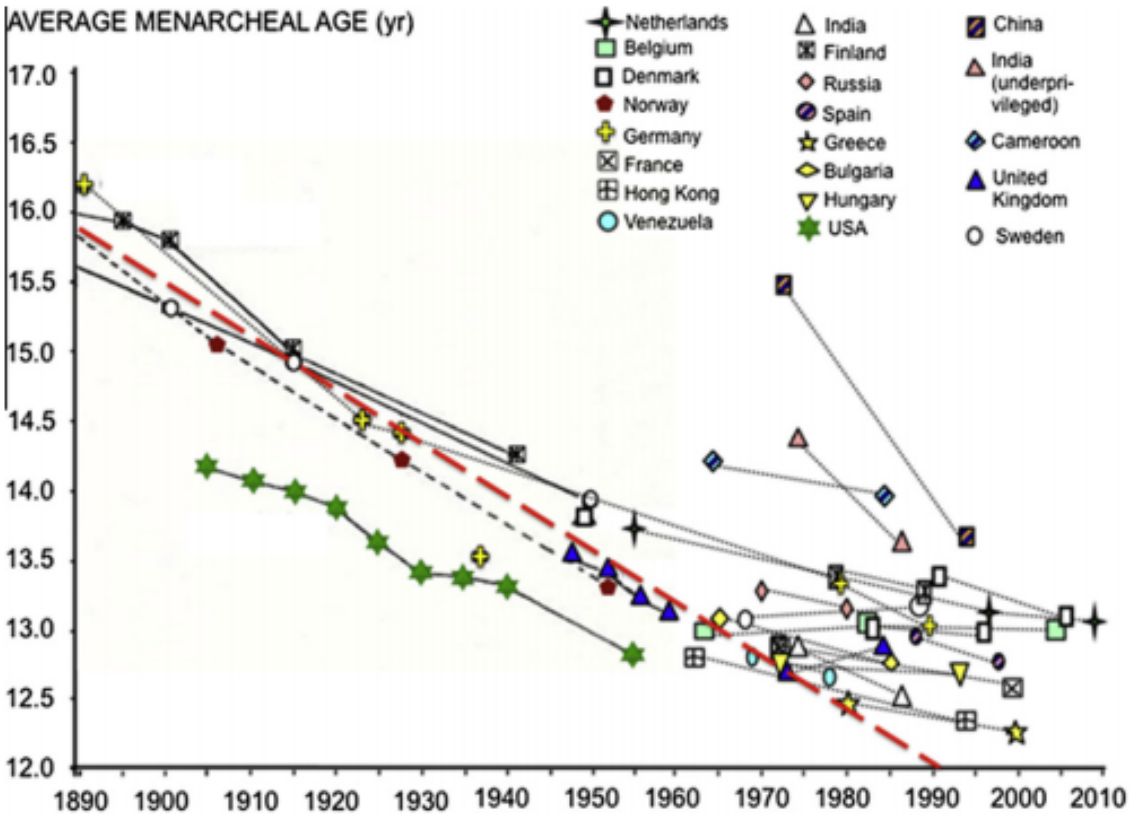


Figure 1.4.1 - Evolution of average menarcheal age (year) in the USA and Nordic countries between 1890 and 1960 (6) Reproduced with author permission.

Much has been postulated about this observed secular trend towards an earlier age of pubertal onset in the developed world (23). Nutritional changes clearly have an important role, as shown by the positive correlation between age at puberty onset and childhood body size, particularly in girls (24). Lower age of both B2 development and menarche has been consistently associated with increased body mass (15). Higher body mass index (BMI) values were seen in early maturers and lower average BMI in late maturers in both white and Afro-Caribbean girls (13). He et al demonstrated in a large dataset (n=3650) that one BMI unit increase between the ages of 2 and 8 years is associated with a 0.11 year advancement in the timing of puberty, as measured by peak height velocity, in both genders (25). In

contrast, under-nutrition in females, for example in chronic disease or anorexia nervosa, can result in delay in both the onset and tempo of puberty (26).

However, in boys the data is less consistent with some studies having noted an earlier onset of puberty with greater adiposity and some with a later onset. In particular, more European studies have noted the former trend, whilst US studies have more often shown the latter (27). Recent data from the USA has shown a far more complex relationship between fat mass and pubertal timing, with overweight status being associated with earlier pubertal onset but obesity associated with later onset. These effects also varied between ethnic groups (28). Thus, one hypothesis in boys is that greater BMI leads to earlier pubertal timing up to the threshold at which obesity occurs. Obesity may lead to later pubertal timing due to suppression of the HPG axis or via adiposity leading to excess aromatase activity and thus increased conversion of testosterone to oestrogen in boys.

This relationship between fat mass and pubertal timing is mediated, at least in part, through the permissive actions of the metabolic hormone leptin, a key regulator of body mass, produced from white adipose tissue (WAT, Fig. 1.1) (29). Leptin levels rise in the serum in early female puberty (30) and are required for normal reproduction. Humans and mice lacking leptin (Lep ob/ob) or the leptin receptor (LepR db/db) fail to complete puberty and are infertile (31). However, the action of leptin in influencing GnRH secretion is not entirely clear-cut. In males, leptin levels decrease during puberty. Leptin does not act directly on GnRH neurons as they do not express the LepR. Leptin appears to indirectly regulate GnRH neurons via its action on the hypothalamus, via cells that are afferent to GnRH neurons such as LEPR-expressing GABA neurons from the arcuate nucleus (32), or via cells that interact morphologically with them, at least in part via the action of nitric oxide (which is required

for its action) (33), and via kisspeptin/neuropeptide Y neurons (34, 35). These upstream regulators of GnRH secretion are discussed in more detail below (section 1.5.2).

Ghrelin and other gut-derived peptides may also form part of the mechanism by which energy homeostasis regulates reproductive development (36). Both low birth weight and prematurity are associated with earlier onset of puberty (37, 38), particularly in those children with rapid increase in length or weight in the first two years of life (39). It remains unclear, however, if childhood obesity, insulin resistance, excess androgens or other factors may explain this association (40, 41).

Despite this weight of evidence, additional data points to a downwards trend in the age of puberty onset that is independent of BMI (42). Moreover, whilst an ongoing strong secular trend towards earlier attainment of B2 has been recognized, the age of menarche in recent years, at least in Northern European studies, has not declined to the same extent (43). Indeed, as detailed above, some studies suggest that over the last decade the age of menarche and of completion of puberty in males in some populations has become skewed towards later ages (6). These data may imply that the increase in fat mass alone cannot explain this secular trend, and suggest a role for factors that have an oestrogen-like effect, without central activation of the HPG axis (19).

The effect of possible endocrine-disrupting chemicals (EDCs) on the timing of puberty has been an ongoing concern (6, 44). Polybrominated biphenyls, bisphenol A, atrazine (herbicides) and phthalates, amongst others, have been suggested as possible EDCs responsible for contributing to this observed trend (45). For example, children migrating for international adoption and formerly exposed to the oestrogenic insecticide DTT in their country of origin displayed early or precocious pubertal timing (46). However, a clear

mechanism of action for EDCs through the early initiation of the pulsatility of GnRH from the hypothalamus has not been conclusively demonstrated. Studies are complicated by the likely differing and possibly divergent influence of different doses and mixtures of EDCs, and differing effects depending on age and length of exposure (6, 46-48).

Epigenetic regulators are potential mediators of the effects of the environment on the hypothalamic regulation of puberty. However, whilst experimental data from rats gives evidence for changes in histone acetylation and gene methylation leading to altered gene expression during puberty, the link between environmental factors and epigenetic control of puberty via the hypothalamus has not been established. Whilst the window of opportunity for the effects of EDC exposure was historically considered to be in the late pre-pubertal period, evidence of fetal and neonatal origin of changes in pubertal timing counters this dogma. Prenatal exposure in boys to EDCs such as phthalates is associated with reduced masculinisation of genital structures (49). Epigenetic changes during fetal life are a potential mechanism for the effects of EDCs in utero (6). Recent evidence suggests that effects of EDCs may persist in pregnant rats in not only their unborn fetus but into the next generation as well (50).

Despite the demonstrated importance of environmental factors, genetic influence on the timing of puberty is clearly fundamental. Whilst the timing of pubertal onset varies within and between different populations, it is a highly heritable trait. The timing of sexual maturation is highly correlated within families and in twin studies, suggesting strong genetic determinants (51). Previous epidemiological studies and genetic approaches estimate that 60-80% of the variation in pubertal onset is under genetic regulation (18, 52, 53). However, despite this strong heritability, little is known about the genetic control of human puberty;

either in the normal population or in cases of disturbed pubertal timing (5). A lack of clear understanding of the genetic factors that control and trigger the onset of puberty is an important barrier to unpicking the role of body mass, EDCs and other environmental cues in the secular trend towards pubertal precocity (54).

1.5 Central control of puberty via GnRH and its upstream pathways

1.5.1 Development of the GnRH neuronal network

The embryonic migration of GnRH neurons from nose to hypothalamus is key for the development of the neuroendocrine pathways that allow normal pubertal development (55). Despite great progress in the understanding of this migratory process, the origin of GnRH neuronal precursors remains unclear. Hypotheses as to their derivation have included nasal placode cells, adenohypophyseal precursors or neural crest-derived cells. In mammals, GnRH1 expression can be detected pre-implantation in the blastocyst and morula stages, but are first detected in the developing head within the nasal placode. GnRH neurons develop in the nasal placode as a cell population distinct from the developing olfactory epithelium (56). Whilst, GnRH cells share some markers in common with neural crest cells, evidence for a direct lineage had been difficult to establish (55). However, data from Cre-lox lineage tracing has led to the current consensus that GnRH neurons are of mixed lineage, with around 30% being neural crest derived and the remainder being of placodal ectoderm origin (57).

The pathway along which the GnRH neurons travel is along and within a scaffold formed by, at different stages, olfactory, vomeronasal and terminal nerves 3 and 4 (1). Cells begin their

migration in the nasal compartment in or around the vomeronasal organ, cross the cribriform plate and penetrate the brain in close proximity to the olfactory bulbs, and finally migrate apposed to a subset of vomeronasal nerves that diverge caudally into the basal forebrain. At the end of their journey, GnRH neurons dissociate from guiding axons to disperse into their final positions of the septo-hypothalamic region.

This whole process of migration involves no more than a few hundred neurons per hemisphere in mouse (several thousand in primate or human) (1). The absolute number of GnRH neurons required for pubertal development is not known, but there appears to be a degree of redundancy in the system. Rodent studies suggest that around 12% of the GnRH neuron population is sufficient for pulsatile gonadotropin secretion and puberty onset, whereas between 12 and 34% are required for cyclical control in adult female mice (58). In addition, adult Reeler mice have significantly fewer GnRH neurons in the hypothalamus and display a phenotype of delayed pubertal maturation and low fertility (59).

GnRH neurons are known to have receptors for at least 20 neurotransmitters. Migratory GnRH neurons receive a plethora of guidance and movement-inducing messages during this journey, which are likely to be distinct depending on the stage of their migration (Fig 1.5.1.1) (60). Signals may act directly or indirectly through the extension of olfactory axons, as disruption of the nerve tract 'scaffolds' themselves can disrupt GnRH migration (61). The molecular signals involved include those controlling cell-to-cell interactions (membrane receptors e.g. neuropilin-2 (62), adhesion molecules e.g. N-CAM (63), extracellular matrix molecules e.g. heparin sulphotransferases (64)), cytokines (e.g. LIF (65), HGF (66)), transcription factors (e.g. Ebf2 (67)) as well as both chemo-attractants and

chemo-repellents (e.g. Reelin (59)). Gradients of chemokines (e.g. SDF1, also known as CXCL12 (68)) may be particularly important for promoting movement of GnRH neurons (60).

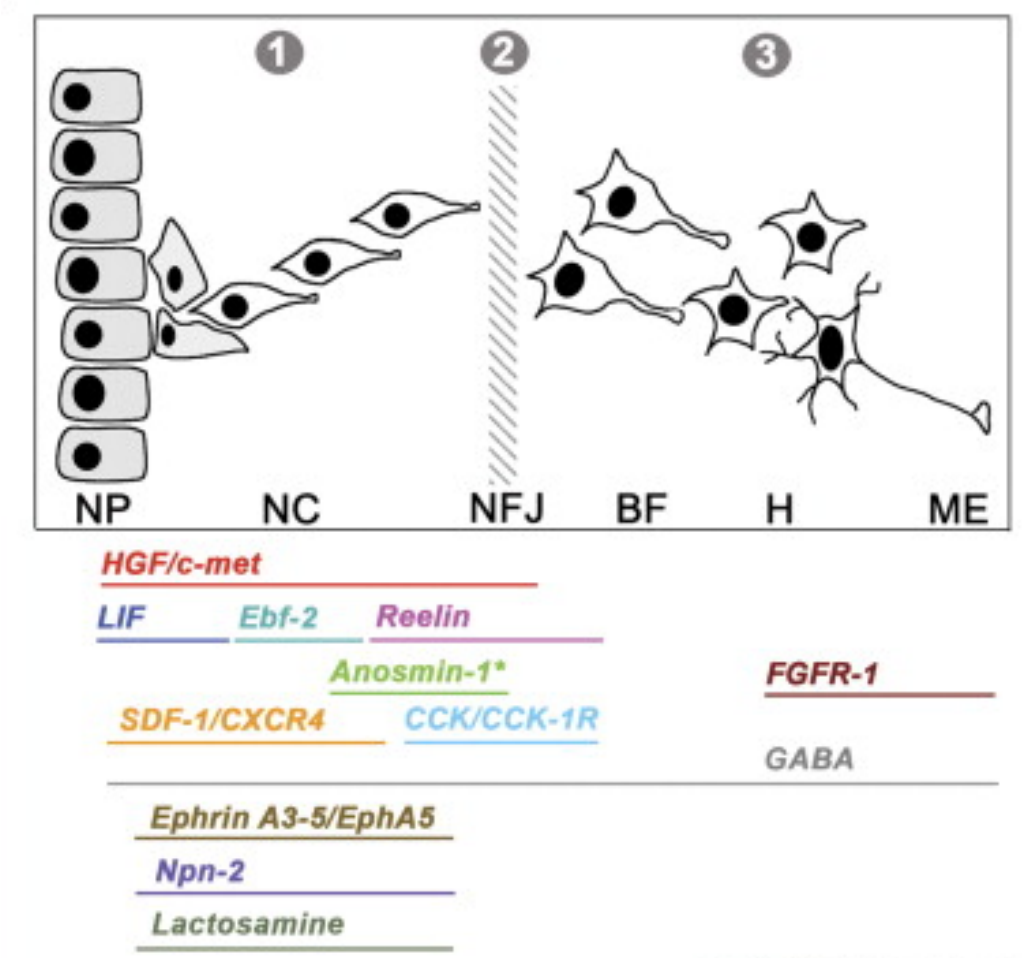


Figure 1.5.1.1 – Factors that affect the migration of gonadotropin-releasing hormone (GnRH) neurons through the three compartments (from Cariboni et al (1), reproduced with author permission). The illustration shows the movement of GnRH neurons from their origin in the nasal placode (NP), through the nasal compartment (NC), and their deflection at the level of the nasal-forebrain junction (NFJ) as they progress toward the basal forebrain (BF). Their migration finally terminates in the hypothalamus (H) from where they project to the median eminence (ME). Factors that have been shown to affect GnRH neurons at different stages of their journey are shown below.

ANOS1, previously known as *KAL1*, encodes anosmin-1, an extracellular matrix protein that regulates axonal path finding and cellular adhesion. Anosmin-1 promotes branching of olfactory bulb neurons (69). Subjects with *ANOS1* loss-of-function mutations have arrest of both GnRH neurons and olfactory bulb neurons at the cribriform plate (70). It is not yet clear whether the effects of anosmin-1 are limited to the development of olfactory neurons, or whether it has an additional chemotactic influence on GnRH neurons (71). Although no mouse model is available, fish and nematode studies and *in vitro* work have further elucidated the role of *ANOS1* (72).

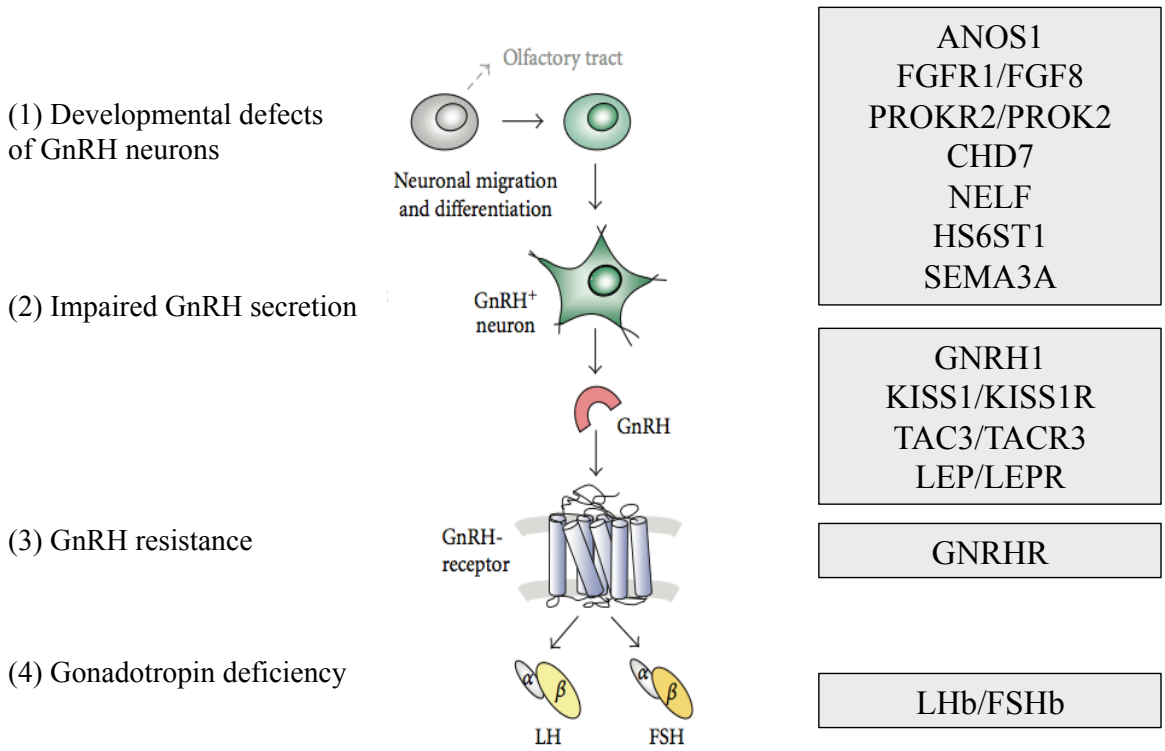


Figure 1.5.1.2 – **Mutations in single genes at many levels of the HPG axis can cause HH** (adapted from Karges B et al (2))

Nasal embryonic LHRH factor (NELF) also appears to have an important role as a common guidance molecule for olfactory axon projections and subsequently, either directly or indirectly, in the migration of GnRH neurons (73, 74).

Perturbations of this migratory journey have been repeatedly shown to lead to hypogonadotropic hypogonadism (HH) in humans and in animal models (75). In addition, survival and differentiation of GnRH neurons upon arrival at their destination is vital (Fig. 1.5.1.2). Mature GnRH neurons extend axons to the median eminence from which they release GnRH into the pituitary portal system. FGFR1 signalling has been shown to be important for this process of axon extension, with reduced projections to the median eminence in transgenic mice expressing a dominant negative FGF receptor (dnFGFR) in GNRH1 neurons (76).

In addition, interactions between these genetic regulators also appear to be important. It has been proposed that Anosmin-1 finely tunes FGFR1 signalling during olfactory and GnRH neuronal development, in an heparan-sulphate dependent manner. Although extracellular matrix components such as heparin sulphotransferases had been identified as potential regulators of axon guidance, *HS6ST1* has only recently been specifically linked to GnRH neuronal development. Using a *C. elegans* model, Tornberg et al (77) provided evidence that *HS6ST1* may act to regulate neural branching in concert with *ANOS1* and *FGFR1.*

1.5.2 Upstream control of GnRH neuronal function

GnRH release is coordinated by a balancing act between a number of inhibitory and excitatory neuronal and glial inputs (Fig 1.5.2) (78). Retrograde tracing studies in mice have

shown that GnRH neurons are subject to a complex neuronal network of inputs from many regions of the brain including hypothalamic nuclei, the brainstem, limbic system, basal ganglia and motor and sensory circuits (79). Inhibitory inputs are primarily from GABAergic and opiatergic neurons, whilst glutamate and kisspeptin are the central excitatory neuronal signals. Glial cells additionally facilitate GnRH secretion via growth-factor derived signalling (80).

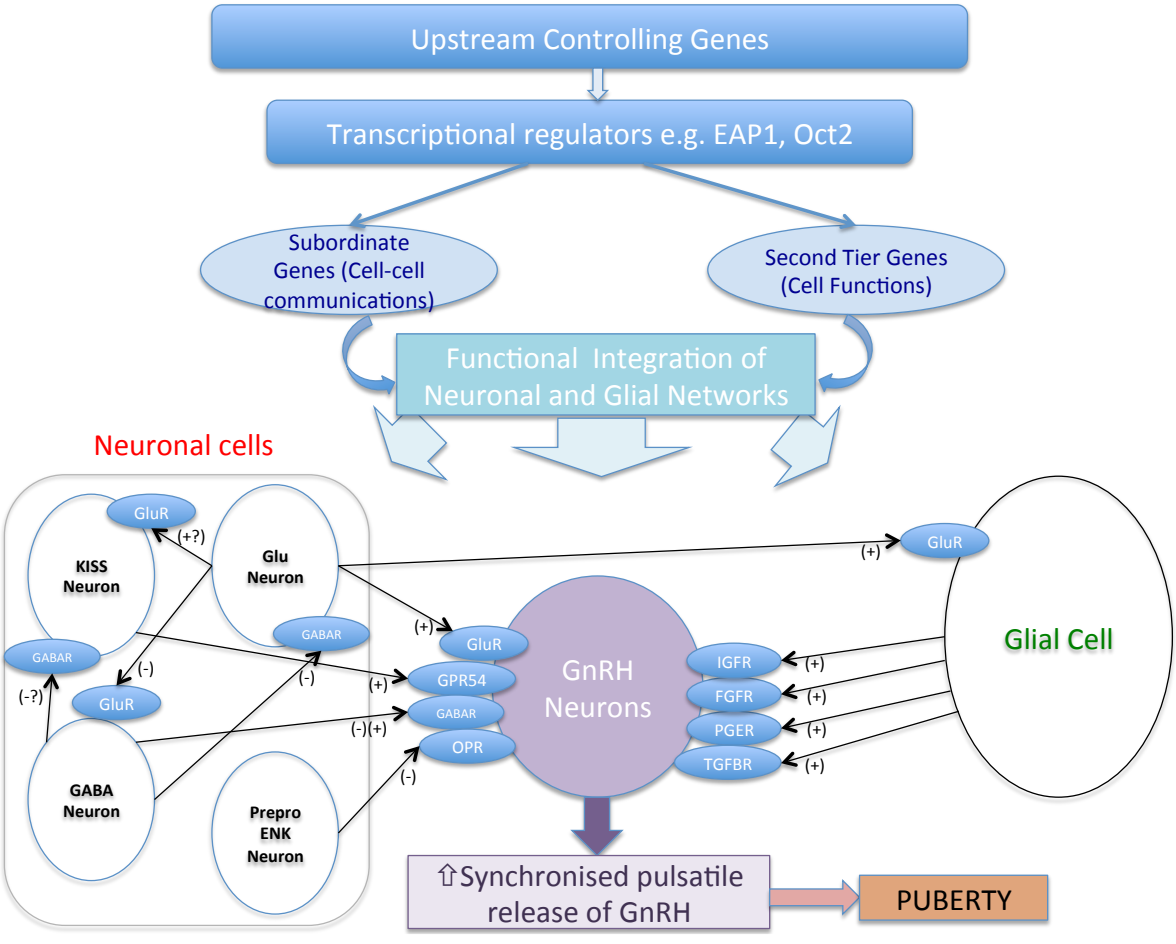


Figure 1.5.2 – Genetic regulators in the transsynaptic and glial control of GnRH neurons during puberty, adapted from (78).

The onset of puberty is triggered by a decline in these inhibitory signals and amplification of the excitatory inputs, leading to increased frequency and amplitude of GnRH pulses. In female rats glutamine synthase is downregulated and glutamate dehydrogenase becomes more abundant in the hypothalamus at puberty, both leading to increased availability of glutamate (81). Glutamate antagonists are potent stimulators of GnRH secretion and administration in prepubertal primates can stimulate the onset of puberty (82).

Kisspeptin, an excitatory neuropeptide, was identified as a vital permissive factor in puberty onset by the discovery of patients with GnRH deficiency with loss of function mutations in the *KISS1* receptor, GPR54 (83, 84). Mice with knockout of *gpr54* were simultaneously discovered to be infertile despite anatomically normal GnRH neurons and normal hypothalamic GnRH levels (84). Their phenotype can be rescued by exogenous delivery of GnRH. *Kiss1* knockout mice also have a phenotype consistent with normosmic GnRH deficiency. However, to date only very rarely have human mutations in *KISS1* have been found in patients with delayed or absent puberty (85), although two potential activating mutations were identified in patients with central precocious puberty (86).

Kisspeptin signals directly to GnRH neurons to control pulsatile GnRH release. Kisspeptin is upregulated in both primates and mice in the peri-pubertal period, and its administration in pre-pubertal rodents advances the onset of puberty (87). Kisspeptin also appears to be downregulated in functional amenorrhoea, suggesting its role as a mediator of the action of environmental factors such nutritional status and emotional wellbeing on puberty and reproductive capacity. Additionally, kisspeptin has been shown to be an important neuroendocrine regulator of ovulation (88).

Kiss1 neurons have been anatomically mapped in rodents (although as yet not definitively in humans) to reside within the hypothalamus within the arcuate nucleus (ARC)/infundibular region, but also as a second population in the anteroventral periventricular (AVPV) region (89). These two populations appear to have distinct roles. AVPV Kiss1 neurons exhibit clear sexual dimorphism, with female rodents at puberty having a much greater number of Kiss1 neurons in this area. Kisspeptin levels in the AVPV of female rodents are amplified in response to peripheral feedback via oestrogen mediated signalling, resulting in a positive feedback loop. In contrast, peripheral oestrogens have a primarily inhibitory effect on Kiss1 neurons in the ARC (89). Thus, kisspeptin signalling is an important element of both positive and negative feedback loops in the HPG axis. However, whilst kisspeptin has been identified as a pivotal upstream regulator of GnRH neurons, whether kisspeptin is the key factor in triggering the onset of puberty remains unclear.

An additional excitatory neuropeptide, neurokinin B, has been implicated in the upstream control of GnRH secretion. Identification of this pathway was also via discovery of loss-of-function mutations in *TAC3*, encoding neurokinin B, and its receptor *TACR3*, in patients with normosmic HH and pubertal failure (90, 91). Both *KISS1* and *TAC3* are expressed by neurons in the ARC of the hypothalamus that project to GnRH neurons, and expression there is downregulated by oestrogen (92). However, evidence for the effects of neurokinin B administration have been somewhat conflicting. Whilst central administration of neurokinin B agonists failed to stimulate GnRH release in rodents, and *Tacr3* knockout mice have grossly normal fertility (93, 94), primate studies have suggested that neurokinin B can act to stimulate GnRH release via kisspeptin signalling (95). Additionally, neurokinin B is expressed

more widely in the central nervous system than kisspeptin, suggesting some differences in the roles of these two neuropeptides in the control of GnRH secretion.

The inhibitory role of GABAergic neurotransmission in restraining the initiation of puberty has been clearly shown in primates (96), but is more ambiguous in rodents. GABAergic signalling pathways are likely to be important in the stress-induced suppression of LH. Opioid peptides provide important inhibitory inputs. Dynorphin, an opioid peptide that is co-expressed with kisspeptin and neurokinin-B in so-called KNDy neurons inhibits the release of GnRH, and together these peptides are currently believed to play a fundamental role in the GnRH pulse generator. Other neurons regulate GnRH including *RFamide-related peptide gene (RFRP)*, the mammalian orthologue of the avian peptide *gonadotrophin-inhibiting hormone (GnIH)*, which has been identified as a further inhibitory regulator of GnRH neuronal activity in mice (97).

Glial inputs appear to be predominantly facilitatory and consist of growth factors and small diffusible molecules, including TGFβ1, IGF-1 and neuregulins, that directly or indirectly stimulate GnRH secretion (98). Firstly, glial cells in the median eminence regulate GnRH secretion by production of growth factors, acting via receptors with tyrosine kinase activity. FGF signalling is required for GnRH neurons to reach their final destination in the hypothalamus (76), as well as for GnRH neuronal differentiation and survival (99). Additionally, GnRH neuron secretory activity is facilitated by IGF-1 and by members of the epidermal growth factor family such as neuregulin 1β (80, 100). Secondly, plastic rearrangements of glia-GnRH neuron adhesiveness, mediated by soluble molecules such as neuronal cell adhesion molecule (NCAM) and synaptic cell adhesion molecule (SynCAM),

coordinate the controlled delivery of GnRH to the portal vasculature (98), a process that is also subject to sex steroid regulation (101).

However, whilst GnRH pulsatility is clearly the central driver of pubertal onset and aberrant GnRH neuronal development and function leads to HH, our collective understanding of the upstream neurocircuitry regulating GnRH neurons is less well established. Data pointing to hypothalamic regulation via a hierarchical network of genes (Fig. 1.5.2) has mainly come from a systems biology approach and animal models, with little data from human subjects (102). Candidate transcriptional regulators that have been identified via these approaches include *Oct-2*, *TTF-1* and *EAP1*. *Oct-2* is a transcriptional regulator of the POU-domain family of homeobox-containing genes. *Oct-2* mRNA is upregulated in the hypothalamus in juvenile rodents, blockage of Oct-2 synthesis delays age at first ovulation and hypothalamic lesions which induce precocious puberty e.g. hamartomas activate Oct-2 expression (103). *TTF-1* (thyroid transcription factor-1) is another homeobox gene that enhances GnRH expression. *TTF-1* expression is increased in pubertal rhesus monkeys (104). *EAP1* mRNA levels also increase in the hypothalamus of primates and rodents during puberty, *EAP1* transactivates the *GnRH* promoter, and EAP1 knockdown with siRNA caused DP and disrupted estrous cyclicity in a rodent model (105).

1.6 Hypothalamus-Pituitary-Gonadal axis activity

The HPG axis is already functional in the late fetal and neonatal period (Fig 1.6.1). During gestation, there is an increase in GnRH content, which reaches its peak at 34–38 wks in the male fetus (106). At mid-gestation, there is a striking rise of circulating gonadotropin levels

in both male and female fetuses, which fall to low levels at birth. This change in gonadotropin secretion results from the development of a negative feedback system through sex steroids, as well as from the development of inhibiting influences from the CNS on GnRH neurons.

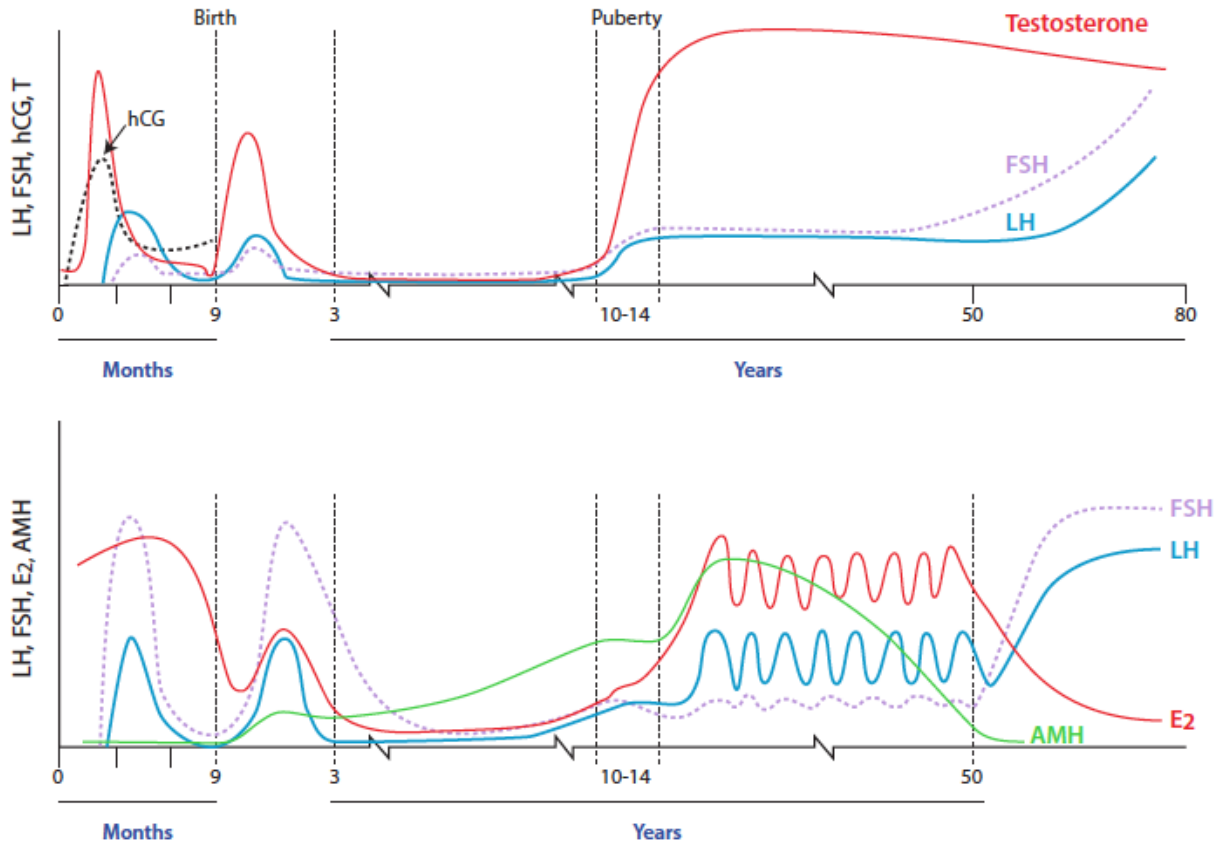


Figure 1.6.1 – The HPG axis during fetal and postnatal life. Circulating concentrations of gonadotropins (hCG, FSH, LH) and testosterone during the life-span of a male (top panel), and gonadotropins, estradiol, and anti-Müllerian hormone during the lifespan of a female (lower panel). From (107).

LH and FSH secretion rise during the first month after birth, probably because the negative feedback effect of placental oestrogens is withdrawn. LH is secreted in pulses during this postnatal period (108), termed the 'mini-puberty'. After this postnatal activity the HPG axis then becomes dormant in children between the age of 2 and 8-9 years (2). This suppression is not absolute, as LH pulsatility is detectable during this stage using ultrasensitive assays, but pulses are infrequent and of low amplitude (109). Development of the clinical features of puberty is initiated by the reactivation of the HPG axis after this period of relative quiescence.

During this reactivation of the axis there is a gradual development of a dynamic interplay between the central production of GnRH and gonadotropins, and gonadal sex steroid production, with progressive maturation of negative and positive feedback loops. The central suppressant drive from the CNS gradually abates and intensifying positive feedback results from the increasing size of, and sex steroid production by, the gonads. Gonadal contribution to the inhibition of the hypothalamic-pituitary system occurs later, becoming operative only at mid-puberty, and eventually becomes dominant over the central inhibitory feedback drive (Fig 1.6.2). Both mean LH and FSH levels increase through pubertal development, although LH rises to a greater extent, probably due to differences in feedback mechanisms for these two hormones (110). These rises are due to both an increase in basal levels of LH and FSH, and a greater number and amplitude of LH peaks.

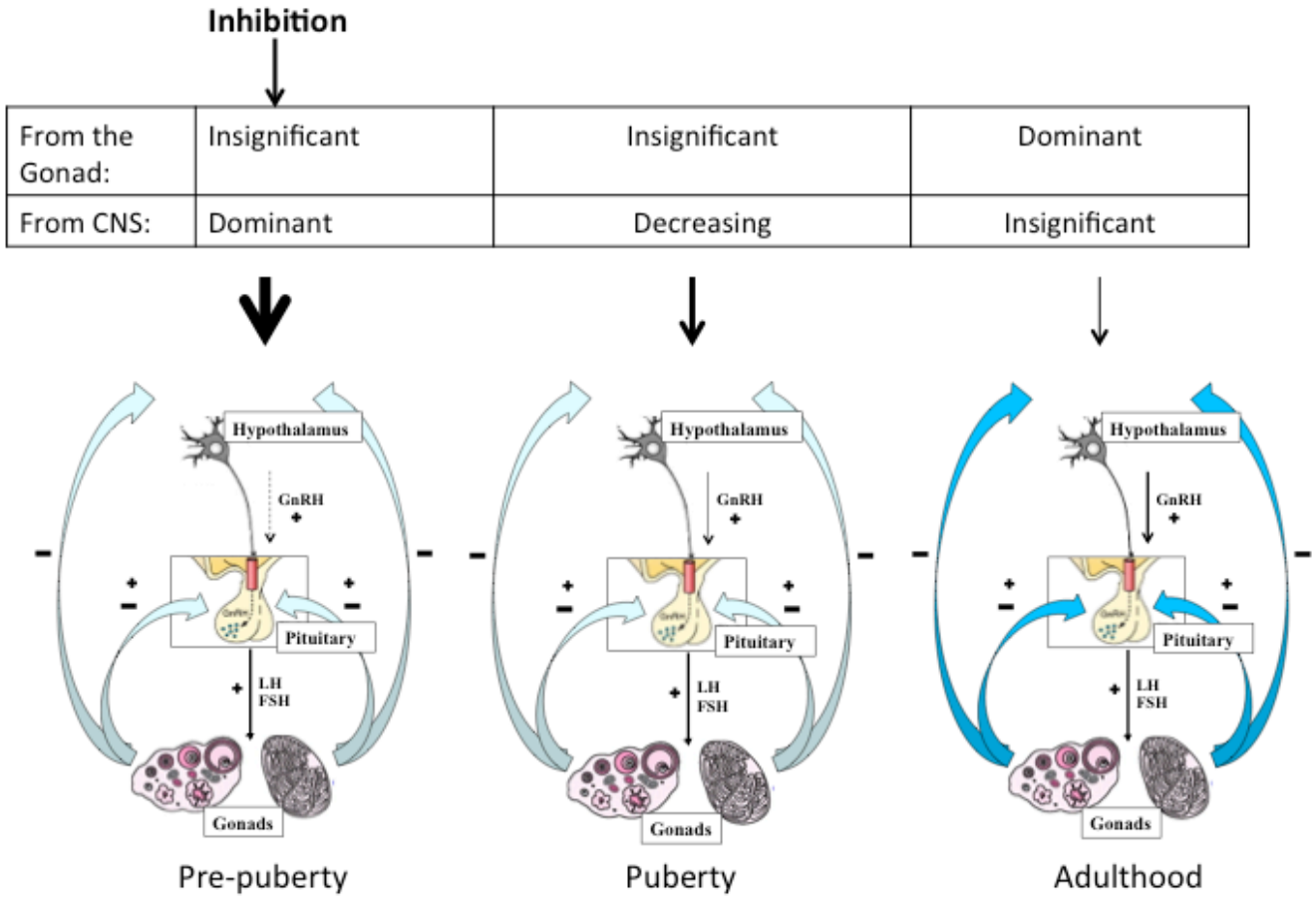


Figure 1.6.2 - Inhibitory regulation of the hypothalamic-pituitary axis. This inhibitory regulation develops gradually after puberty onset, with reduction in the central suppressant drive from the CNS and increasing strength of feedback from the gonads.

1.7 Sex Steroids and Inhibin B

LH controls sex steroid production by upregulating expression of the steroidogenic enzymes and by controlling metabolic activity of the steroid-producing Leydig cells. LH may also stimulate Leydig cell proliferation and differentiation (111). During puberty, plasma testosterone levels increase dramatically. Table 1.7 summarizes the levels of testosterone at various developmental stages, with respective testis sizes. The pubertal increase in testis size results primarily from an increased number of proliferating and differentiating germ cells and, to a lesser extent, an increase in Sertoli cells (112). In early and mid-puberty there

is a pronounced diurnal rhythm with a morning peak in measureable testosterone, but this is less pronounced in later puberty and declines gradually with age (113), probably due to decreased day-night ratios of gonadotropins (114).

Inhibin B is a heterodimeric glycoprotein produced in males exclusively by the testis, primarily in the pre-pubertal testis by the Sertoli cells. Several studies show that serum inhibin-B levels in children change in concert with the secretion of gonadotropins (115-117). During the 'mini-puberty' serum inhibin-B levels increase to similar (118, 119) or higher (116) levels than those observed in adolescent boys and adult men. This early inhibin-B secretion is sustained until the age of 18–24 months; thereafter, serum concentrations decline to lower but readily measurable levels (116). Early in puberty, between Tanner stages G1 and G2, serum inhibin-B concentrations again increase and reach peak levels at Tanner stage G2, but then the levels plateau (115, 117).

Pubertal G-stage (Tanner)	Testicular volume (mL)	Plasma testosterone (ng/dL)	Plasma testosterone (nmol/L)
1	<4	<10	<0.3
2	4–8	12–69	0.4-2.4
3	8–10	60–275	2.1-9.5
4	10–20	142–515	4.9-17.9
5	20-25	319-775	11.1-26.9

Table 1.7: Pubertal Stages (According to Tanner) With Respective Testis Volumes and Plasma Testosterone Concentrations. Values adapted from (120)

1.8 Disturbances of Pubertal Timing

Disturbances of puberty encompass an important group of pathologies within the field of Paediatric Endocrinology. Firstly, they are common, affecting over 4% of adolescents. In addition, abnormal timing of pubertal development is associated with adverse health and psychosocial outcomes (121-123). This has importance both for the individual, but also has a potential major impact on public health, especially in view of the secular trend towards an earlier age of puberty onset (44). Early puberty in particular is associated with adverse health outcomes, including breast (124), and endometrial cancer (125), obesity (126), type 2 diabetes (123), cardiovascular disease (127, 128), short stature (129) and even increased mortality (127). Until recently it had not been clearly shown that late pubertal timing was also associated with adverse health outcomes. However, data from the UK Biobank study from both men and women has demonstrated that DP also has profound impacts on health in later life (130). Late menarche (15-19 yrs) was associated with significantly increased risk of early menopause, osteoporosis, malabsorption, low intelligence, asthma and poor overall health. In addition, increased risk was found for cervical cancer, angina, DVT, myocardial infarction, high cholesterol, still birth and depression. Later voice breaking in men was also significantly associated with anxiety disorders, chronic fatigue syndrome, depression, asthma and poor overall health. Many previous studies have shown that patients with DP are at risk of short adult height (131-134), decreased bone mineral density (135, 136) and psychological distress (137, 138).

1.8.1 Delayed puberty

Self-limited DP, also named constitutional delay of growth and puberty (CDGP), represents the commonest cause of DP in both sexes. Up to 65% of boys presenting with DP have self-limited DP (139). Self-limited DP represents the extreme end of normal pubertal timing, and is defined as the absence of testicular enlargement in boys or breast development in girls at an age that is 2 to 2.5 standard deviations (SD) later than the population mean (5). In addition, self-limited DP may also encompass older children with delayed pubertal progression, a diagnosis that is aided by the use of puberty normograms (Fig. 1.3.1 and Fig. 1.3.2) (10). The absence of pathological medical history, signs and symptoms, and positive family history of DP in one or both of the parents suggests a diagnosis of self-limited DP, but before making the diagnosis, significant pathological conditions (e.g. central nervous system tumours) should be excluded (Fig. 1.8.1).

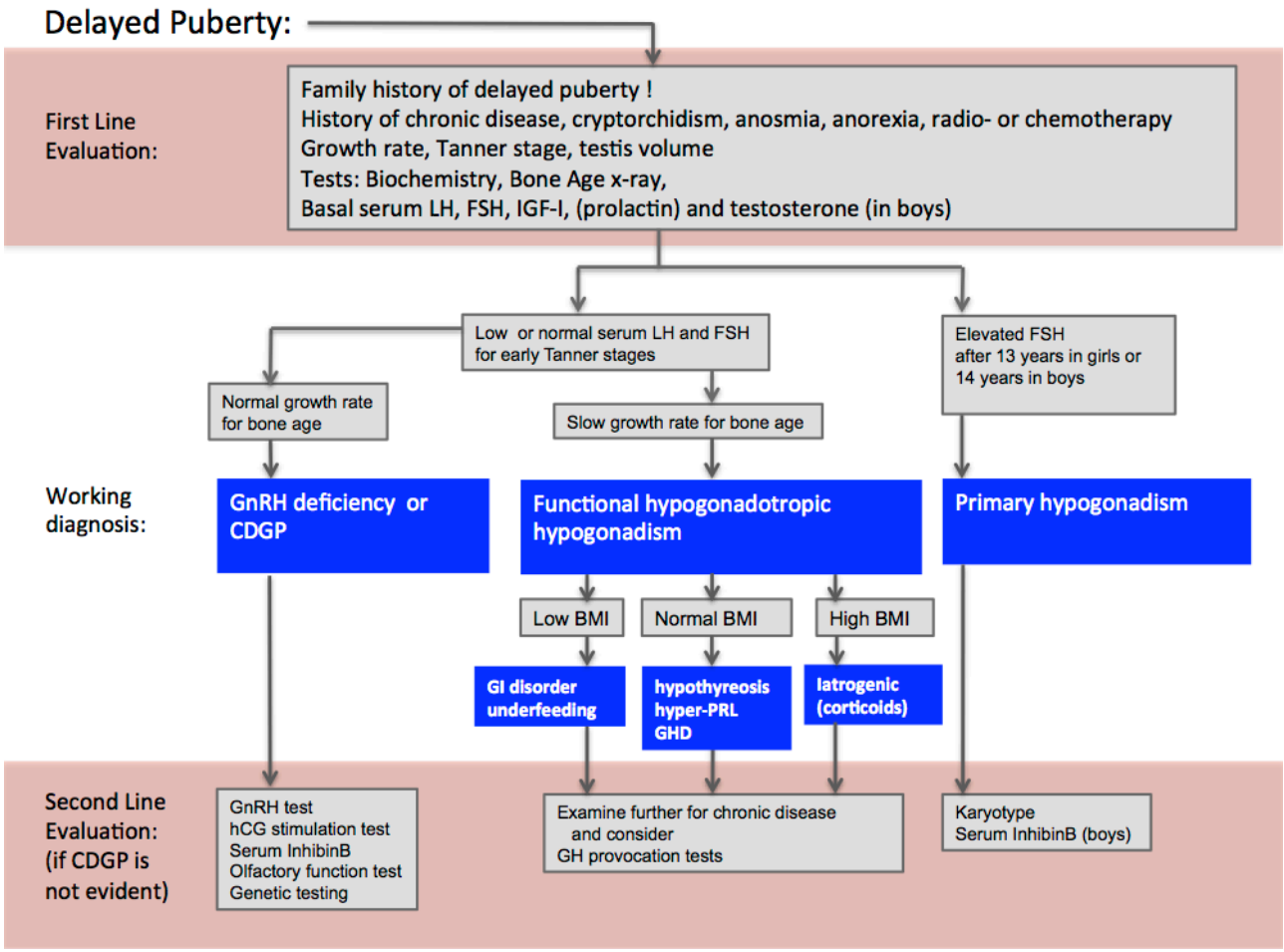


Figure 1.8.1 - Algorithm for the evaluation of a boy with delayed puberty. CDGP - constitutional delay of growth and puberty, GI - gastrointestinal, GH - growth hormone, GHD - GH deficiency, PRL - prolactin, IGF-1 - insulin-like growth factor 1. Figure adapted from (5)

There are three main differential diagnosis of self-limited DP (Table 1.8.1) (5, 139): hypergonadotropic hypogonadism, with primary gonadal failure leading to elevated gonadotropin levels due to lack of negative feedback (found in approximately 7% of male patients and 26% of female patients with DP); functional hypogonadotropic hypogonadism (HH), where late pubertal development is due to maturational delay in the HPG axis secondary to chronic disease (found in 19-20%), malnourishment, excessive exercise, psychological or emotional stress; and permanent HH, characterized by low LH and FSH levels (9% of boys and up to 20% of girls).

	Hypergonadotropic Hypogonadism	Hypogonadotropic Hypogonadism	Functional Hypogonadotropic Hypogonadism
Common Causes:	Klinefelter Syndrome Gonadal dysgenesis Chemotherapy/ Radiation Therapy	CNS Tumours/Infiltrative Diseases Isolated Hypogonadotropic Hypogonadism Kallmann syndrome Combined Pituitary Hormone Deficiency Chemotherapy/Radiation Therapy	Inflammatory Bowel Disease Coeliac Disease Anorexia Nervosa Hypothyroidism Excessive Exercise Cystic Fibrosis

Table 1.8.1: Differential Diagnoses of Male Self-Limited Delayed Puberty. Table modified from (5).

In self-limited DP, probably because of low oestrogen concentration for chronological age (but not for bone age), growth hormone (GH) secretion is functionally and temporally impaired for age, and when this functional GH deficiency lasts for a sufficiently long period of time, it may also have an impact on adult height. After the onset of puberty or the initiation of appropriate oestrogen treatment, growth velocity and GH secretion return to normal. Additionally, in about half of subjects with constitutional delay there is delayed maturation during early childhood, and consequently they are shorter than their peers even before the mean age for the pubertal growth spurt (10). By contrast, patients with congenital HH have steady linear growth during childhood and only become short for their age with absence of the pubertal growth spurt.

It has been shown that those self-limited DP subjects who also have poor growth in childhood may not fully exploit their genetic height potential, resulting in an adult height below their mid-parental target height (134, 140-142), with an average loss of 4.2cm if untreated (142). However, the majority of the patients seen at Paediatric endocrinology clinics are short as their condition is compounded by two independent factors: short genetic height potential and DP. Other studies showed only a negligible difference in final height, even in DP subjects who have received no intervention (131, 143-148). This may imply a pathophysiological mechanism additional to lack of sex steroids contributing to the growth phenotype in some patients with DP, but not in others (148).

In self-limited DP adrenarche may also occur later than usual, in contrast to the normal age of adrenarche in patients with isolated HH (5). Bone age in self-limited DP is retarded compared to chronological age, but the developmental milestones are achieved at a normal bone age; that is, onset of signs of pubertal development by the bone age of 13.5 years in boys and by the bone age of 12 years in girls, as well as menarche by the bone age of 13 years. Gonadotropin and sex steroid concentrations increase in concert with the development of the bone age. Thus, all stages of pubertal development occur at an age later than usual.

Investigation of the differential diagnosis of the two conditions may involve a number of physiological and stimulation tests, including assessment of LH pulsatility by frequent sampling (109), prolactin response to provocation (149), gonadotropin response to GnRH (150, 151), testosterone response to hCG (152-154) and first morning-voided urine FSH and LH (155). Most recently, a single measurement of inhibin B <35 pg/mL in prepubertal boys has been shown to help discriminate HH from self-limited DP with high sensitivity (156), and

a combination of inhibin B with AMH measurement to discriminate with high sensitivity and specificity; however, INSL3, growth and adrenal markers were not found to be helpful markers (157). Treatment of isolated DP involves expectant observation or short courses of low dose sex steroid supplementation.

1.8.2 GnRH deficiency

At the extreme end of the spectrum of DP are conditions of GnRH deficiency including HH, with complete failure to enter puberty. A picture of isolated HH with no associated anatomical or functional defect in the HPG axis occurs in 1-10 cases per 100,000 births. Permanent HH is due to congenital hypothalamic or pituitary disorders, or an acquired central dysfunction. Both GnRH and gonadotropin deficiency can be caused by various genetic or developmental defects of the hypothalamus or by destructive lesions such as tumours, inflammatory processes, vascular lesions, and trauma (158).

Tumours causing DP most commonly interfere with GnRH synthesis or secretion, and deficiency of other pituitary hormones is common. Craniopharyngiomas are the most common neoplasm causing hypothalamic-pituitary dysfunction. At presentation, the most common symptoms are headache, visual disturbances, short stature, DP, polyuria, and polydipsia. Langerhans' cell histiocytosis (LCH), also called Hand–Schuüller–Christian disease or histiocytosis X, may also occasionally present with DP. Germinomas are the most common extrasellar tumours to cause DP, although these tumours are a rarity among primary CNS tumours. Polydipsia, polyuria, and visual disturbances are the most common symptoms associated with these tumours, followed by arrested growth and DP.

Treatment of CNS tumours, leukaemia, or neoplasms with cranial irradiation may also result in gradual development of hypothalamic-pituitary failure. GH deficiency is the most common component of the radiation-induced hormone disorder, but gonadotropin deficiency also occurs when the radiation dose is high enough. Development of radiation-induced hypothalamic-pituitary failure may take from 1 year to several years to ensue.

Because of different causes and incomplete penetrance, there is a wide spectrum of phenotypes of HH, ranging from complete HH with lack of pubertal development to a partial hypogonadism with an arrest of pubertal development, and even reversible HH in some patients, either spontaneously or post treatment (159, 160). Despite recent advances, with over twenty genes linked to this disorder identified, the pathophysiological basis of HH in the majority of individuals remains unclear (2). The condition may be due to failure of development of GnRH neurons, lack of activation of GnRH secretion or disrupted GnRH signalling (Fig. 1.8.2).

Kallmann syndrome (KS) is the most common form of isolated HH, accounting for 60% of cases. Hypogonadism is due to GnRH deficiency, and the other components of the syndrome include anosmia or hyposmia due to hypoplasia of the olfactory bulbs and occasionally cleft lip and palate, unilateral renal agenesis, short metacarpals, sensorineural hearing loss, synkinesia and colour-blindness (161). About 10% of KS patients have mutations in the *ANOS1* (also called *KAL1*) gene on chromosome Xp22.3 (162, 163). As discussed in section 1.5.1, this gene encodes an extracellular matrix protein that regulates axonal path finding and cellular adhesion (164). Defects in this gene cause failure of fetal GnRH neurosecretory neurons, alongside olfactory axons, to co-migrate from the olfactory

placode to the medio–basal hypothalamus, causing hypoplasia of the olfactory sulci and absent or ectopic mature GnRH neurons (70).

In congenital idiopathic HH the diagnosis is typically made during the second or third decade of life. Common presenting signs are delayed onset of puberty, poorly developed secondary sexual characteristics, eunuchoid body proportions, or infertility. In some cases, the diagnosis can be suspected before the age of pubertal onset and, importantly, if this suspicion arises in the first six months of life it can be confirmed on the basis of low testosterone and gonadotropin levels, indicating an absence of the normal 'mini-puberty'.

The presence or absence of "red flag" features remain the strongest discriminator between DP and HH. These red flags include cryptorchidism or micropenis, indicating a lack of prior 'mini-puberty', or the presence of the other components of KS that include anosmia or hyposmia due to hypoplasia of the olfactory bulbs and additional features e.g. cleft lip and palate as described above.

It may be very difficult to distinguish clinically between the conditions of DP and congenital idiopathic HH in the teenage years. In HH, gonadotropin responses to GnRH stimulation may be subnormal, but because of the functional hypogonadotropism in self-limited DP, the differential diagnosis between these two disorders may be difficult. In addition, self-limited DP and HH may share an underlying pathophysiology, at least in some cases. There remains no definitive test to accurately discriminate between the two diagnoses (165).

More complex and involved management is required in patients with hypogonadism to achieve both development of secondary sexual characteristics and to maximize the potential for fertility. Clearly genetic diagnosis of patients in clinic would represent a huge

leap forward in patient care and a likely significant economic advantage. While presently whole exome sequencing in individuals presenting with delayed or incomplete pubertal development is only a reasonable option in a research setting, one might envisage a time where quick and easy diagnosis via an establish gene panel might be available in the clinic setting for patients with both GnRH deficiency and self-limiting DP.

1.8.3 Hypergonadotropic Hypogonadism

Hypergonadotropic hypogonadism, with primary gonadal failure leading to elevated gonadotropin levels due to lack of negative feedback is found in approximately 7% of male patients with DP, and up to a quarter of female patients (139).

The commonest condition underlying hypergonadotropic hypogonadism in males is Klinefelter syndrome (47, XXY), with a prevalence of 1 in 667 live births. The majority of those affected will enter puberty spontaneously at a normal age (166), although DP may be seen in those with a more complex karyotype (48, XXYY, 48, XXXY, 49, XXXXY). Sex steroid replacement is therefore not normally required for these patients at the start of puberty, but testosterone levels become increasingly deficient by Tanner stages 4-5, possibly as a result of secondary regression. However, only 10% of boys aged 10-14yrs with Klinefelter syndrome have been diagnosed and many patients only come to the attention of an endocrinologist in later adulthood (167). The management of anorchic young men, secondary to congenital absence, vanishing testis syndrome or failed orchidopexy, for induction and maintenance of puberty is similar to that in boys with hypogonadotropic hypogonadism.

The syndrome of gonadal dysgenesis (Turner syndrome) is the most common form of hypergonadotropic hypogonadism in females, affecting about 1 in 2500 live-born girls. About half of girls with Turner syndrome have the 45,X karyotype, but about 99% of fetuses with this karyotype abort spontaneously, and in 1 of 15 spontaneous abortions the fetus has the 45,X karyotype (168). Ovarian insufficiency is also apparent at birth, as evidenced by high gonadotropin concentrations during the neonatal period. During childhood, with the development of the CNS-mediated inhibition of GnRH secretion, gonadotropin levels decrease to near normal levels, but by 10 years of age they usually are elevated again. The Müllerian structures (uterus and fallopian tubes) are present but remain infantile if the ovarian failure is not adequately treated with hormone replacement therapy.

Sexual infantilism is one of the most common clinical findings in girls with Turner syndrome. More than 90% have gonadal failure. It is important to remember, however, that up to 30% of girls will undergo spontaneous pubertal development and that 2 to 5% will have spontaneous menses and may have the potential to achieve pregnancy without medical intervention. Pubertal development may be delayed and, in most patients, is followed by progressive ovarian failure. Most patients are small for their gestational age at birth, and the slow growth rate is apparent after 3 years of age. Most girls fail to have a pubertal growth spurt due to insufficient estrogen production in the ovaries. The mean adult height is approximately 143 to 146 cm, depending on both parental heights and the overall height of the child's ethnic population.

Other causes of premature ovarian insufficiency include (169): pure gonadal dysgenesis (phenotypic females with no pubertal development and the 46,XX or 46,XY karyotype without detectable chromosomal abnormalities); X isochromosome, where abnormal

chromosome division results in duplication of identical chromosome arms, most commonly of the long (q) arm (these patients have streak gonads and a similar phenotype to Turner syndrome); post irradiation or chemotherapy (170); autoimmune ovarian failure (171); genetic mutations in the FSH receptor or post-receptor defects in the gonadotropin receptor signalling pathways (172), microdeletions on the short and long arm of the X chromosome, and mutations in several other genes including *POF1*, *POF2*, *DIAPH2*, *FOXL2* and *BMF15*.

1.8.4 Early puberty

Precocious puberty in girls is commonly defined as the onset of B2 before 8 years of age, and as the onset of Tanner stage G2 before the age of 9.0-9.5 years in boys (173). This is still the consensus amongst European clinicians despite the observed trend towards a decreasing age of breast development in girls in the developed world (174, 175). However, this female age cut-off is based on original data in white British girls (4), and more recent data shows that by this definition a significant minority of Afro-Caribbean girls in the US will be classified as having precocious puberty, with B2 by the age of 8 years (13, 176). Premature development of pubic hair has also been shown in several studies to be more common in Afro-Caribbean children. Some centres have argued for a redefining of the age cut-off for precocious puberty in girls to prevent over-investigation and treatment (177). However, concerns remain that pathological causes of central precocious puberty (CPP) in 7-8 year old girls will be missed if the definition is revised (178).

Precocious onset of puberty is approximately five times more common in girls than boys (179). The prevalence of CPP in girls was found to be 0.2%, but only 0.05% in boys, over a

nine year period in one European series (180). The underlying reasons for this gender difference are not clear. However, an underlying abnormality is found far more commonly in boys, suggesting that many cases of female precocious puberty may represent the end of the normal spectrum without underlying pathology. Additionally, it appears that the female HPG axis may be more sensitive to environmental factors such as increased fat mass than the male (26). This is seen also in populations with functional hypogonadism due to weight loss or excessive exercise, where more women than men are affected (181). However, a small trend towards earlier testicular development in boys with increased BMI has also been found (20). Possibly related to this, an increase in boys with idiopathic precocious puberty without an organic cause has been observed recently (182). One study showed that leptin levels, an important permissive mediator in the influence of body mass on pubertal timing, are similar in pre-pubertal boys and girls. However, whilst leptin levels rise in girls at puberty, associated with an increase in fat mass, they fall in boys at puberty (30).

Precocious puberty is categorised into central (gonadotropin-dependent), involving activation of hypothalamic GnRH pulsatility, or peripheral (gonadotropin-independent), comprising a group of conditions that result from autonomous activation of gonadal hormone production. Diagnosis must differentiate between CPP and peripheral precocious puberty. Puberty progression in CPP will classically be symmetrical, whilst in peripheral precocious puberty there may be asynchronous tanner stage development. It is important to remember, however, that priming of the HPG axis in peripheral precocious puberty may lead in time to CPP.

CPP in girls is frequently found to have no identifiable underlying cause; although this is a less common finding in boys (183). Whilst central precocious puberty (CPP) in girls is more

commonly idiopathic than in boys, it is important to exclude an underlying cause. In both genders, it is more common to identify an organic cause in those children with an earlier age of presentation.

CPP may be due to tumours including astrocytomas, gliomas and germ cell tumours, or other hypothalamic lesions including hamartomas, or congenital abnormalities including arachnoid cysts which may be identified on MRI imaging. All conditions which give rise to damage to the hypothalamic-pituitary axis can represent a risk factor for CPP, including radiotherapy to the hypothalamus or pituitary for paediatric cancer (184), hydrocephalus, birth asphyxia or neurodegenerative disease.

Like self-limited DP, CPP often has a strong familial basis (185). Segregation analysis from one study suggested an autosomal dominant inheritance pattern with incomplete sex-dependent penetrance (185). This study demonstrated a significantly earlier age of maternal menarche and a more advanced pubertal staging at presentation in the familial cohort as compared to sporadic CPP. Despite this few genes have been identified as causal in CPP. A heterozygous activating mutation in the *KISS1R* gene has been identified in one Brazilian girl with CPP (186). Rare variants in the *KISS1* gene have been found in a small number of children with CPP (86). However, mutations in these genes have not been found in other cohorts with familial CPP. Most recently, the imprinted gene *MKRN3*, which is involved in ubiquitination and cell signalling, was identified as an important gatekeeper in pubertal onset. Unlike *KISS1* and *KISS1R*, mutations in this gene in familial CPP appear to be relatively common in both genders, identified in 33-46% of probands as compared to 3-6% of patients with sporadic CPP (187-189), However, this has not been verified in all populations (190). Mutations in *MKRN3* appear to be associated with an earlier age of puberty onset in girls as

compared to boys, in keeping with a female HPG axis being more 'sensitive' to disruption by genetic or environmental factors.

Interestingly, the median age of CPP at 6 years in girls with *MKRN3* mutations was not as extreme as with, for example, the *KISS1* mutations that have been identified. This has led to the suggestion that the action of *MKRN3* may not be required to suppress the activation of the HPG axis in early childhood, but instead is important for regulation of pubertal timing in the later pre-pubertal period (189). Abreu et al demonstrated a prominent reduction in *MRKN3* mRNA expression in the arcuate nucleus of mice immediately before puberty (187). Together this adds weight to the hypothesis that *MRKN3* may be part of the mechanism to release the inhibitory 'brake' at the onset of puberty.

Peripheral, gonadotropin-independent precocious puberty is caused by gonadal tumours, adrenal tumours, exogenous sex steroid exposure and in boys, hCG-producing tumours and male-limited precocious puberty. Male-limited precocious puberty, or familial male precocious puberty, is due to an autosomal dominant inherited activating mutation in the LH receptor which presents in males at a young age with penile and testicular growth, advanced bone age and sexualised behaviour (191).

In addition to these causes, peripheral precocious puberty in may also be seen in McCune-Albright syndrome (MAS), aromatase inhibitor excess and congenital adrenal hyperplasia (CAH) (192, 193). Whilst classical CAH more commonly presents in the newborn period, non-classical CAH can present as precocious pubarche (development of pubic hair) with accelerated bone growth (194). McCune-Albright syndrome (MAS) is a predominantly female cause for peripheral precocious puberty. This condition, involving genetic mosaicism, is caused by somatic cell (postzygotic) mutations in the G3α subunit (*GNAS1* gene) of the G3

protein that activates adenylate cyclase (195). This activating mutation can result in multiple endocrinopathies, but early pubertal development in MAS is common, often at two-three years of age.

True precocious puberty in girls must also be distinguished from premature thelarche, usually with breast development before the age of 3 years, and premature pubarche, with the isolated development of pubic hair. The latter may also occur in boys. These conditions are not usually associated with accelerated growth rate or advancement in bone age.

The presentation of precocious puberty is associated with advanced bone age and tall stature due to a precocious pubertal growth spurt. However, if untreated this will result in premature closure of the epiphyses and consequently ultimately adult height is compromised and will be less than mid-parental target height. It can cause emotional distress both to the child and family due to early development of sexualised behaviour and difficulty coping with premature physical changes. In addition, early age of pubertal onset has been linked to multiple adverse health outcomes, as discussed previously in section 1.8.

1.9 Previous Studies of the Timing of Puberty

Attempts to identify key genetic regulators of the timing of puberty in humans have ranged from genome-wide association studies of age-at-menarche (196, 197) examining pubertal timing in healthy women, to next-generation sequencing (NGS) approaches to identify causal mutations in disease cohorts, with both delayed or absent puberty and precocious puberty.

1.9.1 Population studies

The existence of genetic heterogeneity is supported by several large genome wide association studies (GWAS) (126, 197, 198). The first of many loci associated with age of menarche was the gene *LIN28B (199). LIN28B* is a human ortholog of the gene that controls, through microRNAs, developmental timing in the Caenorhabditis elegans. The lin-28 family regulates the biogenesis of let-7 microRNA (miRNA) family members controlling the timing of developmental events, and in turn, let-7 miRNA controls lin-28 translation. The major allele of the single nucleotide polymorphism (SNP) rs314276 (located in intron 2 of *LIN28B*) was associated with earlier age at menarche and earlier breast development in girls, earlier voice breaking and more advanced pubic hair development in boys, and faster tempo of height growth and shorter adult height in both sexes (197). However, mutations in *LIN28B* have not to date been identified in human patients with DP (200) or precocious puberty (201).

In 2010 a large meta-analysis identified 42 (30 new, 2 previously confirmed and 10 possible) loci for age at menarche. In 2014 this was extended to encompass data from genome-wide and custom-genotyping arrays in up to 182,416 women of European descent from 57 studies (Fig 1.9.1). Evidence ($P< 5 \times 10^{-8}$) for 123 signals at 106 genomic loci was identified. Many of these loci were associated with Tanner staging in both sexes, suggesting this data is applicable to both men and women. Further GWAS data appears to confirm this finding (202, 203).

Importantly, for the first time genes already identified in rare disorders of puberty were identified from a genome-wide association study. These included the imprinted gene *MKRN3*, paternally-inherited mutations in which have been identified as causal in pedigrees

of CPP (187). *MKRN3* is to date only the third gene with mutations identified as causal in pedigrees of central precocious puberty (187), the others being *KISS1 (86)* and its receptor *GPR54 (186)*.

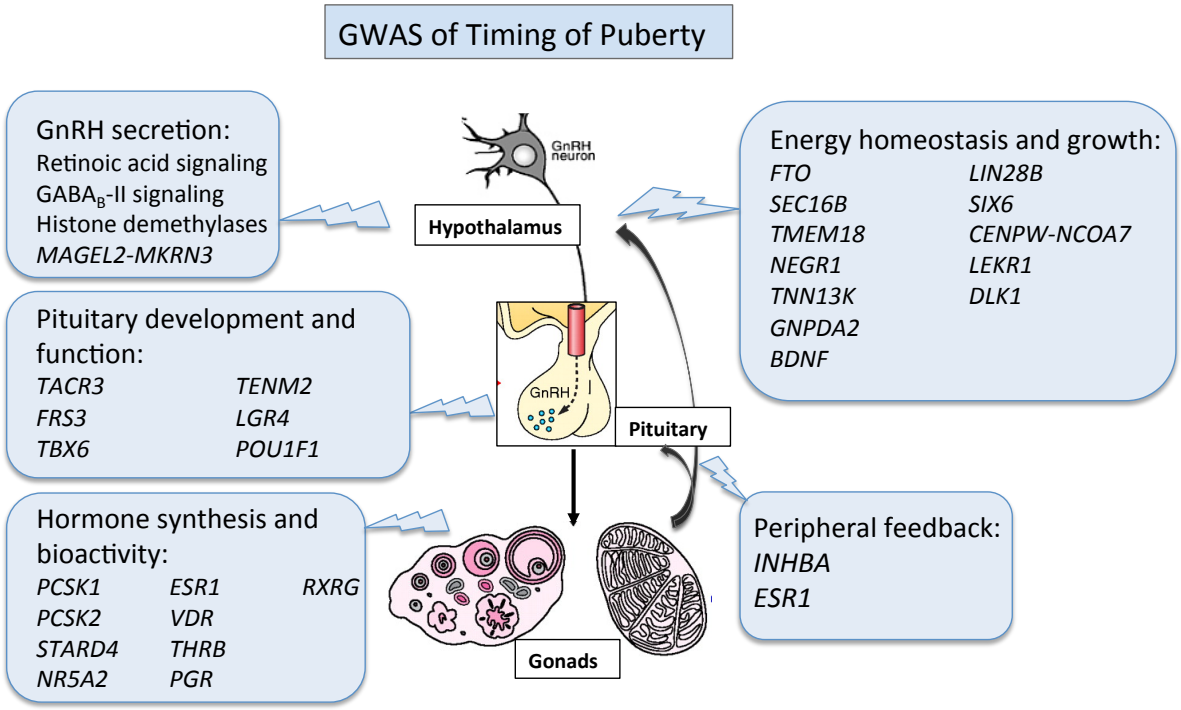


Figure 1.9.1 - Schematic diagram indicating possible roles in the hypothalamic-pituitary-ovarian axis of several of the implicated genes and biological mechanisms for menarche timing (adapted from Perry et al (198)).

Signals were also found near *LEPR-LEPROT*, which encodes the leptin receptor and immediately upstream of *TACR3*, which encodes the neurokinin B receptor. A further variant ~10 kb from *GNRH1* approached genome-wide significance. Two signals were found near *PCSK1* and *PCSK2*, indicating a common function of the type 1 and 2 prohormone convertases in pubertal regulation. Finally, signals in or near several further genes with

relevance to pituitary development and function including *POU1F1, TENM2,* and *FRS3* and signals representing *cis-eQTLs* for *LGR4* and *TBX6,* which both encode enhancers for the pituitary development factor *SOX2*, were identified.

In addition to leptin signalling, the authors found overlap with several genes implicated in body mass index including *FTO*, *SEC16B*, *TMEM18*, and *NEGR1*. As discussed previously, puberty requires a minimum level of energy availability, whilst increased BMI has been shown to be associated with precious onset of puberty (204). However, the molecular mechanisms for this are still unclear. Roles for other genes connected with regulation of body mass have not been clearly demonstrated in puberty. *FTO* (fat mass and obesity associated gene) was the first obesity-susceptibility gene identified through GWAS and continues to be the locus with the largest effect on body mass index and obesity risk (205). *FTO* appears to exert this effect via influence on food intake regulation rather than physical exertion. A recent finding of the importance of *IRX3*, a second gene found at the same GWAS locus, in influencing BMI briefly threw into doubt the primary role of *FTO* (206). However, the weight of evidence behind *FTO*, in particular the results from *FTO*-knockout mice and *in vitro* studies showing that FTO expression is regulated by essential amino acids and that it couples amino acid levels to mTORC1 signalling, have reinforced *FTO* as a major player in the regulation of body mass (205). A novel idea in the analysis of GWAS data is that a number of genes in any one identified region may play an important role in a particular phenotype. Whether such genes may regulate pubertal timing exclusively via impact on body-mass or via other BMI-independent mechanisms is as yet unknown.

Pathway analyses implicated nuclear hormone receptors, particularly those involved in retinoic acid (RA) and gamma-aminobutyric acid-B2 receptor signalling. The active

metabolites of vitamin A, all-trans-RA and 9-cis-RA, have differential effects on GnRH expression and secretion. Other possible mechanisms linking RA signalling to pubertal timing include inhibition of embryonic GnRH neuron migration, and enhancement of steroidogenesis and gonadotropin secretion.

However, together these loci are estimated to explain only 2.71% ($P = 1x\ 10^{-20}$) of the variance of age of menarche (198), suggesting that many of these genetic variants have a low impact in the general population. The authors hypothesise that the genetic architecture of the timing of puberty in healthy subjects involves at least hundreds of common variants.

Additionally, these studies rely on self-recall of the age at menarche, which may result in imprecise data. Research into the major genetic determinants of pubertal timing in the normal population is still in its infancy and is only just beginning to produce feasible candidate genes.

1.9.2 Gene identification in rare disease

Deletion mapping, homozygosity mapping in consanguineous kindreds with multiple affected members, targeted sequencing projects and more recently NGS approaches in patients with congenital HH have led to the identification of genetic mutations in up to 30 separate genes resulting in severely delayed or absent puberty (52, 207-212). These genes include *ANOS1, FGFR1*, *FGF8*, *PROK2* and *PROK2R*, *CHD7* and *NELF*, *GNRH1* and *GNRHR*, *KISS1* and *GPR54*, *TAC3* and *TACR3, LHβ and FSHβ.* These genes are involved in the control of GnRH neuronal migration and differentiation, GnRH secretion or its downstream pathways (Fig 1.8.2) (2). Although these discoveries are critical in enhancing our understanding of the

regulation of these hormonal pathways, approximately 50% of the genetic causes of HH and KS are still unknown (213).

There is a greater or lesser degree of phenotypic variability depending on the gene involved, with nearly all mutations in *ANOS1* leading to KS; whereas loss of function mutations in *FGFR1* have been identified in pedigrees with KS, normosmic HH and DP. Whilst environmental factors may go some way to explaining these variations, it is becoming increasingly clear that gene-gene interactions in HH are an important phenomenon, and strategies to identify such digenic and even oligogenic inheritance in families are developing (214, 215). In such kindreds the pattern of those affected by disease may not conform to classic Mendelian inheritance dogma and bioinformatic filtering pipelines and statistical modelling techniques will require modification in order to identify novel candidates for such gene-gene interactions.

1.9.3 GnRH Resistance and Gonadotropin Deficiency

Loss-of-function mutations within the GnRH receptor are the most frequent cause of autosomal recessive IHH, accounting for 16% to 40% of patients. Mutations have been found within the extracellular, transmembrane and intracellular domains of the receptor leading to impaired GnRH action (216). Treatment of these patients is with gonadotropin therapy.

LH and FSH are glycoprotein hormones encoded by a common α-subunit gene and a specific β-subunit gene. Mutations of the β-subunits of luteinizing hormone (LH) or follicle-stimulating hormone (FSH) are rare causes of HH. Females with inactivating mutations of

LHβ present with onset of normal puberty, but with normal or late menarche followed by infertility due to lack of ovulation (217). Males with inactivating mutations of the *LHβ*-subunit have absent pubertal development with Leydig cell hypoplasia leading to testosterone deficiency and azoospermia. Individuals with inactivating *FSHβ* mutations present with incomplete pubertal development and primary amenorrhea in females and azoospermia in males (218).

1.9.4 Previous genetic findings in self-limited DP

Self-limited DP represents an extreme variant of normal pubertal timing and has repeatedly been shown to cluster in families. Several groups have noted that self-limited DP often segregates in an autosomal dominant pattern, suggesting that inheritance of self-limited DP is conferred by a small number of genes (51, 219).

Linkage analysis on an overlapping sample of our Finish cohort (220) was carried out previously using 383 microsatellite (short-tandem-repeat) markers spanning all autosomes and the X chromosome. This analysis identified a linkage peak at the peri-centromeric region of chromosome 2 with a maximum multipoint heterogeneity logarithm of the odds (HLOD) score of 4.44 after fine mapping. However, no clear candidate genes have as yet been identified within this region, despite follow up sequencing of the complete genomic region of chromosome 2 from 79-124 Mb in the 13 probands and their affected parent who contributed most to the linkage finding. An investigation of low-frequency variation in this region, focusing on coding variation in any gene and regulatory variation in

functionally relevant candidate genes, was unable to detect a susceptibility variant that transmitted along with pubertal delay in these families (221).

In view of the possible overlap between the pathophysiology of DP and conditions of GnRH deficiency, a few studies have examined the contribution of mutations in HH genes to the phenotype of self-limited DP. A homozygous partial loss-of-function mutation in *GNRHR* was found in two brothers, one with self-limited DP and one with idiopathic HH (222), and a further heterozygous mutation was found in one male with self-limited DP (223). A small cohort of 31 patients were analysed for mutations in *GHSR* and 5 patients were found to have point mutations in this gene (224). Of 50 self-limited DP patients investigated for mutations in *TAC3* and *TAC3R*, only one mutation in a single patient was found in the latter gene (225). Recently, variants in several HH genes including *GNRHR*, *TAC3*, *TACR3*, *IL17RD* and *SEMA3A* have been identified by whole exome sequencing in some cases of DP, including self-limited DP (226). Additionally, mutations in *HS6ST1 and FGFR1* have been found a small number of kindreds of HH patients and their relatives with DP (77, 227).

Analysis of self-limited DP families is complicated by the fact that this phenotype represents the tail of a normally distributed trait within the population, so it is expected that variants that govern the inheritance of this condition will also be present in the general population at a low level. Thus, the absence of these variants in population databases cannot be used as an exclusion criterion. Instead, a comparison of prevalence of such variants must be made to identify those that are enriched in patients compared to the general population.

In the majority of patients with DP the neuroendocrine pathophysiology and its genetic regulation remain unclear. Linkage analysis and targeted sequencing strategies appear to

have been superseded by whole exome sequencing strategies to identify novel candidate genes.

1.10 Animal models

1.10.1 Rodents

Rodent models of pubertal and reproductive abnormalities have provided vital information to understand the genetic control of pubertal timing. A minority of models arise from genetic defects occurring spontaneously in breeding colonies, or through fortuitous unplanned outcomes of transgenic mice experiments; for example, the *GNR23*$^{-/-}$ mice colony which have a marked reduction in GnRH neuronal number (58). However, these events occur rarely and instead both chemical and genetic means of modification have been developed for generating new murine models. Whilst chemical mutagenesis is usually random, transgenic models have become increasingly refined and, with the use of conditional models, have allowed knock-out of individual genes in specific systems at particular developmental stages. This approach also allows for generation of mouse models in scenarios where total knock-out of the gene is embryonically lethal.

Assessment of pubertal timing in female rodents is through the onset of vaginal opening, a conventionally accepted external marker of the onset of puberty, and in male rodents by the onset of preputial separation. However, whilst in female rats vaginal opening is associated with full activation of the HPG axis and shortly precedes first ovulation, this association is less definitive in mice. Vaginal opening can also be influenced by exposure to oestrogen containing compounds affecting gonadal development but not reflecting central HPG axis activation.

Measurement of the onset of oestrus cyclicity, and therefore ovulation, by assessing vaginal smear cytology, is a more accurate method of assessing pubertal onset especially in mice, but is also a more involved, time-consuming and complex process (228).

1.10.2 Zebrafish

Zebrafish embryos are both transparent and develop externally, making them an excellent model organism for studying embryogenesis (229). Embryogenesis is complete by 50 hours post fertilization (hpf) with the development of nearly all major vertebrate organ systems firmly established. GnRH neurons develop in a parallel manner to in humans and other mammals, with migration from their origin in the olfactory area to the hypothalamus during the first 5 days post fertilisation (dpf). Zebrafish models, particularly the transgenic (*Tg*) *GnRH3:EGFP* zebrafish line, in which GnRH3 neurons express EGFP, have been utilised to investigate the mechanism of action of several important regulators of GnRH neuronal migration, including SDF-1, Netrin (230) and NELF (74). In contrast, pubertal onset in zebrafish is a relatively late event. Gonadal differentiation is completed around 35 dpf in females and 45 dpf in males. The initiation of puberty in females is at around 45 dpf and is strongly correlated with body size but not age, emphasising the importance of the growth axis in the onset of puberty in zebrafish (231).

1.10.3 Other organisms

Puberty and the timing of pubertal onset has been studied using several other model organisms, ranging from primates to sheep and even the worm Caenorhabditis elegans. Primates have a similar pattern of HPG axis activity to humans with a period of relative quiescence from late infancy until the onset of puberty, and hypothalamic control of the

pituitary-gonadal axis through GnRH was first established in higher primates (232). Primate studies have given insight into the neuroendocrine brake restraining the HPG axis during childhood and to the hormonal mechanisms that result in the release of that brake (233). C. elegans has been used to demonstrate the role of Lin28, as discussed in section 1.9.1.

1.11 Thesis Rationale

Self-limited delayed puberty (DP) segregates in an autosomal dominant pattern and is highly heritable, but the underlying neuroendocrine pathophysiology and genetic regulation are unknown. The limited understanding of the factors that trigger pubertal onset constrain our ability to make accurate diagnoses and optimally manage patients with pubertal disorders. Furthermore, this has hampered attempts to comprehend the population-wide trend towards a more precocious age of pubertal onset in the developed world. Previous studies have suggested that up to 80% of the inheritance of pubertal timing is genetically determined. Identification of these critical alleles has not been feasible by GWAS alone, probably due to their low frequency.

My hypothesis was that discovery of novel genetic regulators of pubertal timing may be achieved through next generation sequencing of subjects with pubertal timing at the tail ends of the normal population distribution, with an extreme phenotype inherited via one or a few genetic variants. Our cohort of pedigrees affected with familial self-limited DP provides a highly valuable resource for this purpose, as it has a comparatively homogeneous genetic composition and is likely to be enriched for low frequency, high or moderate effect alleles.

My first aim was to use an unbiased approach to identify novel genetic factors influencing delayed pubertal onset, through whole exome sequencing of probands and their relatives. Utilising an in-house bioinformatics pipeline designed for this purpose, I aimed to pinpoint mutations in key genes that produce the phenotype of DP in one or more families. Through comprehensive functional characterisation, I attempted to gain an understanding of the role of any identified genetic regulators in the timing of puberty.

Secondly, I aimed to assess the degree of shared genetic regulation between late pubertal timing in patients with self-limited DP, and absent puberty in those with HH. I explored the exome data generated from our patients with self-limited DP for mutations in genes already implicated in conditions of GnRH deficiency. Thirdly, I assessed the degree of shared genetic regulation between late pubertal timing in patients with self-limited DP and the timing of puberty in the general population. I examined the loci identified in previous studies of pubertal timing via GWAS of AAM, for any overlapping genetic candidates identified by the exome sequencing of our cohort. Finally, I aimed to assess the potential importance of regulators of GnRH neuronal development in the pathogenesis of self-limited DP, by utilising microarray data analysing embryonic GnRH neuronal development to explore the results of our exome data from our self-limited DP cohort.

Ultimately, I hoped to provide insights into the pathogenesis of DP and other conditions of aberrant pubertal development. Such discoveries may transform patient management and also identify potential novel therapeutic targets. This work will also offer important insights into the understanding of the control of puberty in the normal population, including the still unanswered question of ‘what triggers puberty?’.

Chapter 2: Materials and Methods

2.1 Patients

DP: The patients selected for this study are taken from a previously collected and described Finnish DP patient cohort, which is highly accurately phenotyped and characterised with the diagnosis based on objective evidence of a delayed pubertal growth spurt rather than self-recall (220). Patients referred with DP to specialist paediatric care in central and southern Finland between 1982 and 2004 were identified. All patients (n=492) met the diagnostic criteria for self-limited DP, defined as the onset of Tanner stage G2 (testicular volume >3 ml) at >13.5yr in boys or Tanner stage B2 at >13.0yr in girls (i.e. two SD later than average pubertal development) (5). Medical history, clinical examination, and routine laboratory tests were reviewed to exclude those with chronic illness. HH, if suspected, was excluded by spontaneous pubertal development at follow-up. 50% of patients from the cohort chose to have pubertal induction via the use of exogenous sex steroids. All of these patients were followed up once off treatment until the point of full pubertal development (Tanner stage G4+ or B4+), to ensure that pubertal development did not arrest off treatment.

Families of the DP patients were invited to participate, with information about medical history and pubertal timing obtained by structured interviews and from archived height measurement records. The criteria for DP in probands' family members were one of three of the following: 1) age at take-off or 2) peak height velocity (PHV) occurring 1.5 SD beyond the mean, i.e. age at take-off exceeding 12.9 and 11.3 yrs, or age at PHV exceeding 14.8 and 12.8 yrs in males and females, or 3) age at attaining adult height more than 18 or 16 yrs, in males and females, respectively (220).

HH: With the aim of identifying whether mutations in *IGSF10* contribute to the pathogenesis of HH, clinical collaborators were identified with existing cohorts of adult patients with idiopathic HH or Kallmann Syndrome. A HH cohort of patients with idiopathic hypogonadotropic hypogonadism (IHH, n=158), Kallmann Syndrome (KS, n=162) and functional HH (hypothalamic amenorrhea, HA) or HA-equivalent (n=14) were collected through 3 coordinating European centres (Newcastle upon Tyne Hospital, Newcastle, UK; Bicêtre University Hospital, Paris, France; Robert Debré Hospital, Paris, France). Patients with IHH all had absent or incomplete puberty at 18 years of age, low or normal serum gonadotropin levels, low serum oestradiol or testosterone levels, otherwise normal anterior pituitary function, and normal results on neuroimaging. KS patients had absent or incomplete puberty at 18 years of age, low or normal serum gonadotropin levels, low serum oestradiol or testosterone levels, otherwise normal anterior pituitary function, and anosmia diagnosed clinically and by radiological evidence of olfactory bulb disruption. Patients with HA or HA-equivalent had a history of spontaneous pubertal development with secondary amenorrhea in women for 6 months or more, low or normal gonadotropin levels, low serum oestradiol or testosterone levels, and one or more predisposing factors. These factors included excessive exercise (>5 hours per week), or other stress, loss of more than 15% of body weight, and either evidence of a subclinical eating disorder, or another cause of dietary restriction (e.g. self-perceived “intolerance” to certain foods). Patients with frank anorexia nervosa or BMI <17kg/m2 were excluded.

2.2 Study Approval

The study protocol was approved by the Ethics Committee for Pediatrics, Adolescent Medicine and Psychiatry, Hospital District of Helsinki and Uusimaa (and extended to encompass Kuopio, Tampere and Turku University Hospitals) (570/E7/2003). UK ethical approval was granted by the London-Chelsea NRES committee (13/LO/0257). The study was conducted in accordance with the guidelines of The Declaration of Helsinki. All patients gave informed written consent prior to inclusion in the study. All mouse experiments were conducted under the animal (scientific procedures) act 1986, project licence PPL 70/8269. The human embryonic and fetal material was provided by the Joint MRC/Wellcome Trust grant# 099175/Z/12/Z Human Developmental Biology Resource (http://hdbr.org).

2.3 Genetic analysis

2.3.1 DNA extraction from whole blood

The illustra Nucleon Genomic DNA Extraction Kit (GE Healthcare, Buckinghamshire, UK) was used to extract genomic DNA from patient whole blood. 3-5 ml of patient blood was collected in sodium EDTA tubes. Blood stored at 4°C was extracted within 24 hours of collection, or alternatively samples were stored at -20°C. The procedure, as per the manufacturer's instructions, involved initial cell lysis, deproteinisation with sodium perchlorate, extraction with chloroform treatment and nucleon resin and finally DNA recovery and DNA washing. Details of the buffers used in the kit can be found in Appendix 5. The DNA pellet was then air-dried for 10 mins before adding 100µl of sterile water. The sample was subsequently rotary mixed for 2 hours to dissolve the DNA.

2.3.2 Whole Exome Sequencing

Initial genetic analysis was performed in 18 probands with DP (male n=16, female n=2) from the most extensive families from our cohort and their affected and unaffected family members. The 18 most extensive families were identified by virtue of the number of affected individuals contained within the related family, presence of unaffected relatives to provide control data and absence of missing phenotypic data. Bilineal inheritance families were excluded from this initial selection due to the added complexity of filtering causal variants in bilineal autosomal inheritance models.

Whole exome sequencing (WES) was performed on DNA extracted from peripheral blood leukocytes of 111 individuals from these 18 families (a total of 76 individuals with DP; male=53, female=23 and 35 controls; male, n=13, female n=22). High quality DNA was required with a concentration of 20 ng/µl to 500 ng/µl and an OD260/280 ratio of greater than 1.8 and OD260/230 ratio of greater than 1.9. Absorbance of the samples was measured using a NanoDrop ND-1000 spectrophotometer at λ= 260 nm/280 nm and 260nm/230nm respectively.

WES samples were prepared as an Illumina sequencing library, and in the second step, the sequencing libraries were enriched using either the Nimblegen V2 (samples processed by Otogenetics Corporation, USA) or the Agilent V4 (samples processed by Otogenetics Corporation, USA or BGI Tech, Hong Kong) exome enrichment microarray. Briefly, DNA was fragmented and a library was created using standard techniques including end repair, A-tailing and paired adaptor ligation. DNA was hybridised to the array, washed to remove un-tethered strands, and then enriched, eluted and amplified. The enriched DNA was subjected to massively parallel sequencing on the Illumina HiSeq 2000.

Downstream analysis of the raw sequencing FASTQ files was conducted by Dr Claudia Cabrera, Centre for Translational Bioinformatics, William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, using the following bioinformatics pipeline (designed by Dr Cabrera and Dr Michael Barnes, also of the Centre for Translational Bioinformatics, William Harvey Research Institute). Paired-end reads were aligned to the UCSC hg19 reference genome using Burrows-Wheeler Alignment tool (BWA-MEM [bwa-0.7.12])(234), and Picard Tools software ([version 1.119] http:// http://broadinstitute.github.io/picard/) was used to sort alignments for each paired-exome sequence. PCR duplicates were marked using Picard MarkDuplicates tool. The Genome Analysis Toolkit [GATK-3.4-46] was used to mark and realign sequences around indels. The Mills Indels and the 1000 genomes datasets were used as reference to mark the regions around the known indels. Mate-pair information was verified and fixed if needed with picard-tools. Quality score recalibration was performed with GATK using dbSNP, Mills, and 1000 genomes as reference resources. Variant calling and joint genotyping using pedigree information was performed using HaplotypeCaller in GVCF mode from GATK. The resulting single nucleotide polymorphism and indel variants were filtered using machine learning approaches with the variant quality score recalibration (VQSR) from GATK. VCF files were annotated using ANNOVAR (http://annovar.openbioinformatics.org, [v20151203]).

I then performed analysis of the called variants using two methods: via Ingenuity Variant Analysis (QIAGEN Redwood City, www.qiagen.com/ingenuity) and subsequently, for verification, via manual filtering of annotated VCF files using Microsoft Excel. Filtering for potential causal variants was carried out using filters for quality control, predicted functional annotation, minor allele frequency (MAF), segregation with trait, variants in multiple

families and biological relevance (Fig. 2.3.2). Quality control included thresholds for call quality, read depth and Phred strand bias. Variants called with a genotype quality (GQ) <50 were excluded. Predicted functional annotation involved prioritizing nonsense, exonic missense, splice site variants, structural or promoter changes, or variants deleterious to a microRNA.

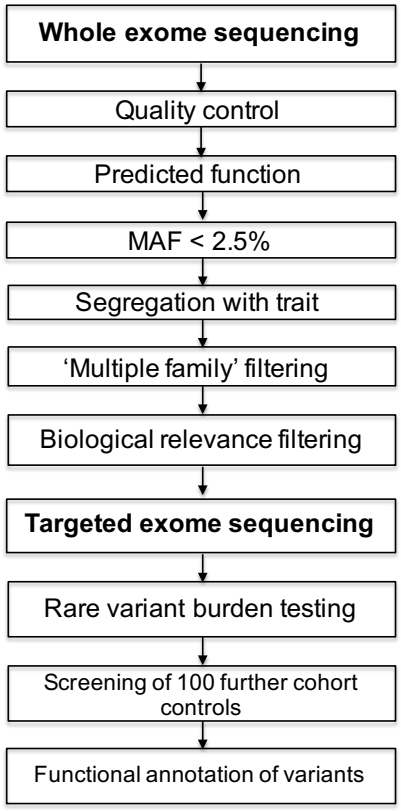


Figure 2.3.2 – Principle components of the whole exome sequencing filtering pipeline. See methods for details of each filtering step.

Filtering by MAF entailed including those variants with minor allele frequency (MAF) <2.5% in the 1000 Genomes database, the NHLBI exome variant server, Sequencing Initiative Suomi (sisu.fimm.fi) and dbSNP databases. Segregation with trait refers to variants present

in ≥ n-1 affected individuals (where n = number of affected individuals in a given pedigree) and not present in more than one unaffected individual being retained. Multiple family filtering involved retaining those variants seen in more than one family from the cohort sequenced, or with different variants in the same gene in more than one family.

Biological relevance filtering allowed prioritization of those remaining variants under 3 criteria:

1) Variants in genes known to be relevant to the phenotype of HH (see Appendix 1). 25 genes in the published literature at the time of searching (May 2012) had been identified with mutations thought to be causal in HH. These 25 'HH genes' were used to identify potential variants from the WES data.

2) Variants in genes in linkage disequilibrium with loci associated with GWAS of age at menarche. To identify variants in known candidate genes that may be causal in self-limited DP, published literature was searched for GWAS studies of the timing of puberty. The 42 loci for age at menarche (AAM) identified by the large meta-analysis by Elks et al (196) were used as the basis for identifying potential candidate genes for the timing of puberty in the general population. All genes in linkage disequilibrium with these GWAS AAM loci (using inclusive limits: D' > 0.8; r^2: no limit) were selected using the Broad institute SNAP tool, (SNP annotation and proxy search, www.broadinstitute.org/mpg/snap/ldsearch.php). Linkage disequilibrium data was calculated using Haploview 4.0, based on phased genotype data from the International HapMap Project and the 1000 Genomes Project. A total of 297 genes were selected using this SNAP tool, and GWAS relevance filtering allowed identification of those remaining variants that lay within these 297 genes. This process was repeated prior to the microarray analysis in Chapter 7, using data from the largest and most recent GWAS of

age at menarche at that time (198). This generated a list of 804 genes in linkage disequilibrium with the GWAS loci, and this list was used for the microarray comparison in section 7.3.

3) Variants in genes with potential biological significance, using the tools Ingenuity Variant Analysis (QIAGEN Redwood City, www.qiagen.com/ingenuity), Genego MetaCore (Thomson Reuters), www.uniprot.org, OMIM and Annovar (235). Data was searched on each gene variant for relevance to biological pathways related to the phenotype of DP, pubertal timing, GnRH biology or central control of the HPG axis.

2.3.3 Targeted exome sequencing

Targeted exome sequencing was performed using a Fluidigm access array assay (www.fluidigm.com). Target-specific primers of the top 28 prioritised candidate genes identified through the previous 7 filtering steps were designed by and purchased from the Fluidigm assay design engine and used to prepare sequencing-ready libraries for NGS. The Fluidigm Access Array utilises an integrated fluidic circuit (IFC) consisting of a microfluidic chip that systematically combines multiple sample inputs with multiple primer inputs to enable large combinations of samples and primer pools. Once samples and primers are loaded and combined, the chip undergoes thermal cycling to amplify regions of interest from the samples.

Upon completion of PCR, the resulting amplicons recovered on a per-sample basis in the IFC were taken for sequencing on the Illumina Miseq. Library preparation and sequencing was carried out by the Genome Centre team at QMUL (www.smd.qmul.ac.uk/gc/). This targeted sequencing was performed in a further 42 families from the same cohort (288 individuals, 178 with DP; male=106, female=69 and 110 controls; male=55, female=58). Downstream

analysis of this NGS data was again conducted by Dr Claudia Cabrera using the same in-house pipeline as for the WES data. Variants returned post targeted exome sequencing were filtered using Microsoft Excel with the same criteria as the WES data in order to identify rare, potentially pathogenic variants: on the basis of quality control, predicted functional annotation and minor allele frequency.

Thereafter, rare variant burden testing was performed for all 28 genes post targeted sequencing. Fisher's exact test was used to compare the prevalence of rare, deleterious variants in our cohort with ethnically-matched samples from the Finnish population, using the ExAC Browser (Exome Aggregation Consortium (ExAC), Cambridge, MA: http://exac.broadinstitute.org, accessed September 2015). For each of the 28 genes, all variants from the ExAC database with minor allele frequency <2.5%, predicted to be deleterious by both Polyphen-2 (236) and SIFT (237), were included in the analysis, with each family in our cohort represented by the proband only. A multiple comparison adjustment was applied to the set of 18 p values post hoc using the Benjamini & Hochberg method (238). Variants identified following this filtering pipeline were confirmed by conventional Sanger sequencing, and screened in a further 100 control subjects (unaffected family members) from our cohort via Sanger sequencing.

2.3.4 Genetic Analysis – HH cohort sequencing

The HH cohort (n=334) was screened via targeted exome sequencing for mutations in *IGSF10* using a Fluidigm array for library preparation followed by Illumina Miseq sequencing, again carried out by the Genome Centre team, QMUL, with the same filtering pipeline used for the whole and previous targeted exome sequencing data. Variants identified were confirmed by conventional Sanger sequencing.

2.4 Polymerase chain reaction (PCR) and sequencing

2.4.1 Oligonucleotide design

The relevant sequence of the gene of interest was established using Ensembl (see URL). Primer sequences for genomic DNA correspond to sequences within the intronic regions whilst primer sequences for cDNA traverse intronic regions. The following guidelines were then used for PCR primer pair design:

- Oligonucleotide length of between 18 to 30 nucleotides
- GC content should be between 40-60%
- Primer pairs should have similar Tm (melting temperature) values, where the Tm can be estimated as follows; Tm = 2°C x (number of A and T residues) + 4°C x (number of G and C residues).
- Complementary sequences within a primer sequence and between pairs should be avoided.
- Runs of 3 or more Gs or Cs should be avoided at the 3' end.
- Complementarity of 2 or 3 bases at the 3' ends of primer pairs should be avoided to reduce primer-dimer formation.

The stock primer solution (Sigma Aldrich, Poole, UK) was stored at -20°C. To avoid repeat freeze-thawing small working aliquots of 10 µM were prepared.

2.4.2 Polymerase Chain Reaction (PCR)

Each variant in genes of interest was amplified by PCR using specific primers (for primer sequences, see specific methods sections). The reaction mixture made up to a volume of 25µl contained the following reagents:

- 100 ng DNA template
- 2.5 µl 10 x Standard Taq polymerase Buffer (10 mM Tris-HCl, 50 mM KCl, 1.5 mM MgCl2, pH8.3) (New England Biolabs, Hertfordshire, UK)
- 0.5 µl, 200 µM each dNTP (dATP, dTTP, dCTP, dGTP; Promega, Southampton, UK, Southampton, UK)
- 200 nM each primer (forward and reverse)
- 0.125 µl Taq polymerase (5000 u/ml) (New England Biolabs, Hertfordshire, UK)
- Nuclease-free water (Sigma Aldrich, Poole, UK) to make up total PCR reaction volume to 25 µl

Cycling conditions, unless otherwise stated, were: 95˚C for 5 mins (1 cycle); 95˚C for 30 seconds, 55˚C for 30 seconds, and 72˚C for 30 seconds (30 cycles); and 72˚C for 5 mins.

PCR products were then separated and visualized using agarose gel electrophoresis. The PCR products were run on 1-2% agarose gel made in 1x TAE buffer (40mM Tris-acetate, 2mM disodium ethylenediaminetetraacetate (Na_2EDTA), pH 8.3; National Diagnostics, UK) supplemented with GelRed (Biotium, Cambridge BioScience, Cambridge, UK, 1:10000 dilution) for visualization, alongside DNA markers (GeneRulerTM DNA Ladder Mix, 0.5 mg DNA/ml, Fermentas). Prior to loading into the gel wells 5 µl of each reaction was mixed with 1 µl of loading dye solution (40% w/v sucrose, 0.25% w/v bromophenol blue or Orange G, 1mmol EDTA pH 8.0). Electrophoresis was carried out at approximately 120 V for 30 minutes or until clear separation of products was obtained. GelRed intercalated into DNA fluoresces under UV light at 300 nm allowing the DNA and therefore the PCR product to be visualized. A Uvitec transilluminator was used to visualize the bands and capture an image of the gel.

The approximate size of the PCR product can be determined by comparison with the DNA markers on the agarose gel.

PCR products were sequenced using the ABI Prism Big Dye sequencing kit and an ABI 3730 automated DNA sequencer (Applied Biosystems, Foster City, CA), carried out by the Genome Centre team, QMUL.

2.4.3 Sanger sequencing analysis

Analysis of sequence chromatograms was carried out using BioEdit (see URL). BioEdit is a freeware biological sequence alignment editor and analysis program for Windows 95/98 NT. The program enables the user to compare the DNA sequence to a reference sequence, enabling identification of differences between 2 or more sequences by eye. This program can therefore be used to detect potential mutations in the chromatograms generated from patient samples.

2.4.4 cDNA production

A human tissue RNA panel was obtained from Agilent Technologies (Stockport, UK) and reverse transcribed to produce a cDNA panel. For cDNA production via reverse transcription, 2 µl of 2 mg/ml RNA in 15 µl RNase-free water (Sigma-Aldrich, Poole, UK) was incubated with 2 µl random primers (hexadeoxynucleotides, 500µl/ml; Promega, Southampton, UK) at 70°C for 5 mins, followed by incubation on ice for 2 mins. A 1 µl aliquot was taken as a pre-reverse transcription negative control which should not contain cDNA or genomic DNA. 1 µl reverse transcriptase enzyme (MMLV-RT enzyme, Promega, Southampton, UK) was added, with 5 µl 5x reverse transcriptase buffer (50mM Tris-HCl (pH8.3), 75 mM KCl, 3 mM MgCl2 and 10 mM DTT; 5x MMLV RT-buffer, Promega, Southampton, UK), 1.25 µl dNTPs (deoxynucleotidetriphosphates ; consisting of 10 mM

dATP, 10 mM dCTP, 10 mM dGTP, 10 mM dTTP; Promega, Southampton, UK) and 0.75 µl RNase inhibitor (RNasin; Promega, Southampton,UK). The entire mixture was incubated at 37°C for 60 mins. The resulting cDNA was used as a template for PCR amplification.

For gel extraction, PCR bands were cut out of the gel and purified using the Qiagen QIAquick Gel extraction kit (Qiagen, Crawley, UK) according to the manufacturers' instructions. The procedure involves silica-membrane-based purification of DNA fragments from gels and details of the buffers used in this kit can be found in Appendix 5. DNA concentration in ng/ml was determined using a NanoDrop ND-1000 spectrophotometer, with the purity of the DNA determined by calculating the ratio of absorbance at 260 nm and 280 nm. A ratio of ~1.8 was considered to be pure DNA. DNA samples were stored at 20°C.

2.5 Cell culture

2.5.1 Cell culture conditions

NIH3T3 cells (sourced from ATCC, Middlesex, UK) and COS7 cells (sourced from Sigma-Aldrich, Poole, UK) were cultured in Dulbecco's Modified Eagle Medium/F12-Ham (1:1) (Sigma-Aldrich, Poole, UK) supplemented with 10% fetal calf serum (FCS; Invitrogen, Paisley, UK) and 1% penicillin/ streptomycin solution (Pen/Strep; Invitrogen, Paisley, UK). All cells were incubated in a humidified incubator at 37°C and 5% CO_2. Stable lines were obtained by application of a selection antibiotic, puromycin at a concentration of 5 µg/ml (Sigma Aldrich, Poole, UK). Cells were checked for mycoplasma contamination (MycoAlert Detection Kit, Lonza) on a monthly basis and were contamination-free.

Passaging and harvesting of cells was carried out as follows: cells were cultured to 90% confluency in a 75 cm2 (T75) flask, cells were washed twice in 5 ml of Phosphate Buffered

Saline (PBS) solution (Sigma Aldrich, Poole, UK) prior to incubation with 3 ml of Trypsin/Ethylenediaminetetraacetic acid (EDTA, Invitrogen, Paisley, UK) until detachment of cells was achieved. Neutralisation of TE buffer was achieved by addition of 7 ml of media and cells were then divided into separate flasks or spun at 1000 rpm for 5 minutes (mins) until a cell pellet was formed and the supernatant was discarded.

Freezing cells for future use: Cell pellets from individual T75 flasks were obtained as described above. A solution of 2.7 ml FCS together with 0.3 ml Dimethylsulfoxide (DMSO; Sigma Aldrich, Poole, UK) was prepared. DMSO acts as a cryoprotectant preventing cell death during the freezing process. The prepared solution was poured over the cell pellet and the pellet resuspended and the suspension divided into 2 vials for freezing. The vials were wrapped in bubble wrap and placed in a -80°C freezer overnight. The following day the vials were placed in liquid nitrogen storage. Following recovery from storage, cell vials were placed in a water bath at 37°C for a few minutes until thawed. 5 ml of media was added to the cell suspension and transferred to a 15 ml falcon tube for centrifugation at 1000 rpm for 5 mins. The cell pellet was subsequently resuspended in fresh media for plating.

2.5.2 Preparation of LB agar plates and LB broth

Lauria Broth (LB) broth (10g LB Broth (Sigma Aldrich, Poole, UK) and 500 ml distilled H2O) and LB Agar (10g LB Broth, 10g Agar (Sigma Aldrich, Poole, UK) and 500 ml distilled H2O) were both autoclaved prior to use. To prepare LB agar plates LB agar was simmered for 10 minutes and allowed to cool. Once cool to touch 50 ml LB agar was poured into a 50 ml falcon tube and 50 µl of 100 mg/ml Ampicillin (Sigma Aldrich, Poole, UK) was added. 25 ml of the above was poured evenly into each plate and allowed to set with a lit Bunsen burner kept in close vicinity to help maintain a sterile environment.

2.5.3 Transformation of chemically competent bacterial cells

NEB 5-alpha Competent E. Coli (High Efficiency) (New England Biolabs, Hertfordshire, UK) were used for high efficiency transformation. Competent cells were stored at -80°C. For use, one tube of NEB 5-alpha Competent E. Coli cells was thawed on ice for 5 mins. The contents of a single tube were then divided into five 1.5ml microcentrifuge tubes for use. 1-5 µl containing 50 ng of plasmid DNA was added to the cell mixture. The tube was flicked carefully 4 to 5 times to mix cells and DNA. The mixture was then placed on ice for 2 mins following which the mixture was heat shocked at exactly 42°C for exactly 30 seconds. The microcentrifuge tube was then placed on ice for 2 mins following which 950 µl of room temperature (RT) SOC Outgrowth Medium (New England Biolabs, Hertfordshire, UK) was added to the mixture. The cell mixture was centrifuged at 4000g for 4 mins following which 850 µl of supernatant was discarded. The remaining 100 µl of the cell mixture was spread onto pre-prepared LB agar plates containing 100 µg/ml ampicillin, inverted and incubated overnight at 37°C for colony formation.

2.5.4 Midi-prep preparation

A single colony was picked from a freshly streaked (overnight incubation) LB agar ampicillin selection plate and used to inoculate a starter culture of 3 ml LB broth containing 100 µg/ml ampicillin. This was incubated for approximately 8 hours at 37°C with vigorous shaking (300 rpm). The starter culture was then diluted 1/500 into 50 ml ampicillin selective LB broth. This was grown at 37°C for 16 hours with vigorous shaking (300 rpm). Bacterial cells were harvested by centrifugation at 4000g for 30 mins at 4°C. The supernatant was discarded and all traces of supernatant removed by inverting the open centrifuge tube until all medium had been drained. Midi-prep preparation was carried out using the Qiagen HI SPEED Plasmid

Midi Kit (Qiagen, Crawley, UK). For details of the composition of each of the buffers see Appendix 5.

2.5.6 Transient Transfection with Lipofectamine-2000

Lipofectamine-2000 reagent (Invitrogen, Paisley, UK) was used for transient transfection. For a standard 6-well plate, 3 µg plasmid DNA was diluted in 100 µl Opti-MEM (Invitrogen, Paisley, UK), mixed gently by pipetting and incubated at RT for 5 minutes. During incubation, a master mix of Opti-MEM (100 µl/well) and Lipofectamine-2000 (5 µl/well) was prepared. The Lipofectamine-2000/ Opti-MEM solution was mixed slowly with the DNA/ Opti-MEM solution, and incubated for a further 20 minutes at RT. Media was removed from wells and replaced with 1.3 ml fresh cell culture media. The solutions of DNA/ Lipofectamine-2000 were added to wells, and cells were incubated overnight at 37°C with 5% CO2. Media was replaced the following morning. Depending on the size of the culture vessel, transfections were scaled up or down as follows:

Culture Vessel	**Opti-MEM (µl)**	**DNA (µg)**		**Lipofectamine-2000 (µl)**	**Opti-MEM (µl)**
6 well plate	100	3		5	100
T25 flask	250	8		15	250
T75 flask	1500	24		40	1500

Table 2.5.6. Transfection with lipofectamine-2000 according to culture vessel type

2.6 Western blotting

2.6.1 Cell lysate preparation

Cells were lysed with RIPA buffer (50 mM Tris-HCL, pH8.0, with 150 mM sodium chloride, 1% IGEPAL CA-630 [NP-40], 0.5% sodium deoxycholate, and 0.1% sodium dodecyl sulphate) (Sigma Aldrich, Poole, UK) supplemented with complete, Mini, EDTA-free Protease Inhibitor Cocktail tablets (Roche Diagnostics Ltd, West Sussex, UK). 150 µl of RIPA buffer was used per well of a 6 well plate, placed on ice for 30 minutes and centrifuged at 15,000g for 12 minutes at 4˚C. The supernatant was subsequently added to an equal volume of reducing Laemmli loading buffer (Sigma Aldrich, Poole, UK) and stored at -80°C for later use. To obtain lysates in non-reducing conditions, the supernatant was added to a non-reducing buffer (250 mM Tris HCl, pH 6.8, with 8% SDS, 40% glycerol, 0.02% bromophenol blue) and stored at -80°C for later use.

2.6.2 SDS-PAGE

Samples were heated to 95-100˚C for 5 mins and 10 µg of protein loaded on pre-cast 4-12% polyacrylamide NuPage BisTris gels (Invitrogen, Paisley, UK). The gels were run in 1 x3 3-(N-morpholino)propanesulfonic acid (MOPS) running buffer (50 mM MOPS, 50 mM Tris, 0.1% SDS, 1 mM EDTA, pH 7.7) for 60 mins at 170 volts. Novex sharp pre-stained protein standard (Life Technologies, Paisley, UK) was loaded alongside the samples.

2.6.3 Immunoblotting

Following size-separation, proteins were transferred for 30 mins at 15 volts to a nitrocellulose membrane by semi-dry transfer using the Trans-Blot SD semi-dry transfer cell (Bio-Rad, Hemel Hempstead, UK). The transfer buffer consisted of 20 mM Tris, 120 mM glycine and 10% methanol. Blots were then blocked with 5 ml of 5% non-fat milk powder

(ASDA, UK) in PBS with 0.05% Tween (PBST). All incubations, unless otherwise stated, were performed for 1 hour at RT in blocking buffer. The membranes were subsequently incubated in 5 ml of 5% milk in PBST containing the appropriate antibody (antibody dilution and incubation period specific to each antibody). Following incubation with the primary antibody, membranes were washed three times in PBST, each wash lasting 20 mins. Membranes were subsequently incubated in appropriate species-specific infra-red secondary antibodies (Licor, Cambridge, UK) at a dilution of 1:5000 for 1 hour, unless otherwise stated. After three 15 min washes in PBST, membranes were imaged using the Odyssey ® Fc Imaging System (Licor).

2.7 Statistics

For rare variant burden testing, Fisher's exact test was used with a multiple comparison adjustment applied post hoc using the Benjamini & Hochberg method (238). A threshold of $p < 0.05$ was taken as significant. Fisher's exact test was also used for analysis of HH cohort allele frequencies. The unpaired t test (2-tailed) was used for statistical analysis of auxological data.

For all experiments, data are expressed as the mean ± SEM. To determine statistical significance, we used the unpaired t test (2-tailed) or, for multiple comparisons, a 1-way ANOVA followed by a Dunnett's post-hoc test. A p value of < 0.05 was considered statistically significant. Statistical analysis was performed using GraphPad Prism4 (GraphPad Software), apart from analysis of mouse pubertal timing in WT versus *FTO* heterozygotes, which was via SPSS (http://www-03.ibm.com/software/products/en/spss-statistics, version 24) to perform linear modelling.

2.8 Cross-Phenotype Analysis

The ReproGen Consortium comprises work from 175 researchers from over 100 research institutes, and has amassed data from GWAS of AAM in the normal population. Through our collaboration with Prof. Ken Ong and Dr John Perry, MRC Epidemiology Unit, University of Cambridge, we have examined their data for variation in our candidate genes. We have tested to what extent any variants in candidate genes reached statistical significance in this AAM data and, therefore, to what extent variants in these genes might influence pubertal timing in the general population. Additionally, through this consortium we have been able to analyse data from the UK Biobank study (www.ukbiobank.ac.uk). This resource gives access to health data from over 500,000 participants that have been sequenced by the UK Biobank Axiom Array (which comprises 820,967 genetic markers). Thus, we can assess particular genotypes, identified via our whole exome sequencing pipeline, for their impact on pubertal timing and other variables including BMI.

2.9 Growth Pattern Analysis

The pattern of pre-pubertal growth in affected individuals carrying pathogenic variants in *IGSF10*, *FTO* or *HS6ST1* was analysed using two parameters: 1) HSDS distance from target height (TH) at the ages of 4 and 8 (girls) or 9 (boys) years (TH formula = 0.791 × mean parental height SDS − 0.147 for girls and 0.886 × mean parental height SDS – 0.071 for boys) and 2) change in height SDS (ΔHSDS) between the ages of 4 and 8/9 years. Normal values for the two parameters, based on data from >70,000 healthy Finnish children, have been previously published (239).

2.10 Functional Annotation of *IGSF10*

2.10.1 *In silico* Analysis

The IGSF10 protein has not previously been modelled by crystallography, so we therefore performed *in silico* analysis. The analyses were carried out by Dr Alessia David, under the supervision of Professor Michael J Sternberg, both at the Centre of Bioinformatics and System Biology, Department of Life Sciences, Imperial College London.

The amino acid sequence was retrieved from UNIPROT database (id Q6WRI0). Homology modelling was used to determine the 3D structure using two high-performing protein structure prediction servers: Phyre2 (240, 241) and I-Tasser (242). These methods construct protein models by detection of homology between a submitted protein sequence and proteins of known structure. They can routinely and confidently detect remote homologous relationships even when sequence identity is <30% and have been shown to be amongst the best systems of their kind in blind trials of structure prediction (241, 243). I-TASSER builds structure models based on the assembly of predicted homologous substructures, using multiple-threading approaches; insight on protein function is derived by matching the predicted models with protein function databases. The signal peptide cleavage site was calculated using SignalP4.1 (244).

The following interactions involved in protein stability were considered: i) salt bridges, defined as at least one pair of atoms on oppositely charged groups within a 4.5 Å distance; ii) hydrogen bonds (H-bond), defined as a donor–acceptor distance ≤2.5 Å and an angle at the acceptor ≥90°; and iii) disulphide bridges (S-S bridge) defined as the side chains of two cysteines at a 3.0 Å distance. The protein electrostatic potential was calculated using the PBEQ program (245), which computes the protein electrostatic potential by solving the

Poisson–Boltzmann equation. Protein structures were visualised using the Pymol visualisation program (http://www.pymol.org/).

2.10.2 Quantitative real-time PCR

A quantitation assay is a real-time PCR assay which measures the amount of nucleic acid during each amplification cycle of the PCR. For this purpose, fluorescent SYBR Green I dye chemistry was used. This uses the SYBR Green I dye which binds to double stranded DNA formed during PCR. During the PCR, DNA polymerase amplifies the target sequence, which creates the PCR products (amplicons). The SYBR Green I dye then binds to each new copy of double stranded DNA. As the PCR progresses more amplicons are created resulting in an increase in fluorescence intensity proportionate to the amount of PCR produced.

RT-qPCR was performed in a 10 μl reaction containing: 2 μl cDNA template, 5 μl 2X $SYBR_{GREEN}$ I Master Mix (KAPA Biosystems), 0.2 μl low ROX (KAPA Biosystems), 0.5 μl primers (10 mM forward + reverse) and 2.3 μl nuclease-free H_2O. Glyceraldehyde-3-phosphate dehydrogenase (*Gapdh*) was used as the endogenous housekeeping gene.

The real-time PCR was performed using an Mx3000 Thermocycler (Stratagene) using the following primers:

	Human-IGSF10	Human-GAPDH
Forward	5’- CAGCAGCATTGTCTTTCCA -3’	5’- GAAGGTGAAGGTCGGAGTC -3’
Reverse	5’- ACCAGTTGGAGTCACTTGGG-3’	5’- GAAGATGGTGATGGGATTTC -3’

	Mouse-Igsf10	Mouse-Gapdh
Forward	5'-CTGGGGAGTCCAATTGCTGT-3'	5'-TGCACCACCAACTGCTTAG-3'
Reverse	5'-GCTGCCTTTGCTGACATC-3'	5'- GGATGCAGGGATGATGTTC -3'

Quantitative RT-PCR was set up in triplicate. After an initial denaturation step of 3 min at 95°C, PCR cycling was performed for 40 cycles of 95°C for 3s, 55°C for 20s and 72°C for 1 sec, followed by 1 cycle of 1 min at 95°C, 55°C for 30s and 95°C for 30s. A no template control (NTC) sample that does not contain template was used to verify amplification quality.

2.10.3 Constructs and protein expression

After failure to identify a successful and specific commercial antibody against IGSF10 to be used for *in vitro* work, a strategy to produce an EGFP-tagged IGSF10 protein for expression in mammalian cells was agreed upon. An Myc-DDK-tagged ORF clone of Homo sapiens IGSF10 (RefSeq NM_178822.3) was purchased as transfection ready DNA (Origene, Rockville, US). However, despite numerous attempts, I was unable to express the full-length IGSF10 protein in mammalian cells, likely due to its large size (7869 base-pairs, 295 kDa). Instead, in order to explore the functional consequences of the 2 N-terminal variants identified in our patients, an N-terminal fragment of the *IGSF10* gene was sub-cloned into a pcDNA-EGFP (Addgene plasmid #13031, see Appendix 3). This N-terminal fragment encoded the first 668 aa of the IGSF10 protein, in view of evidence from previously published work of a putative cleavage site at this position (246). Briefly, following digestion of the Origene vector with appropriate restriction enzymes and gel purification (section 2.5.3), the purified product was ligated with the restriction-digested Addgene plasmid and, after transformation and amplification, verified by sequencing.

The p.Arg156Leu and p.Glu161Lys variants were generated using PCR-mediated mutagenesis (QuikChange II Site-Directed Mutagenesis Kit, Agilent Technologies, Stockport, UK, see Appendix 5) according to the manufacturer's instructions. Briefly, a control reaction was prepared containing: 5 µl of 10× reaction buffer, 2 µl (10 ng) of pWhitescript 4.5-kb control plasmid (5 ng/µl), 1.25 µl (125 ng) of oligonucleotide control primer (100 ng/µl), 1.25 µl (125 ng) of oligonucleotide control primer (100 ng/µl), 1 µl of dNTP mix and 38.5 µl ddH2O (to bring the final reaction volume to 50 µl), after which 1 µl of PfuUltra HF DNA polymerase (2.5 U/µl) was added. A sample reaction was prepared containing 5 µl of 10× reaction buffer, 5µl (50 ng) of dsDNA template, 0.5µl (125 ng) of oligonucleotide primer 1, 0.5µl (125 ng) of oligonucleotide primer 2, 1 µl of dNTP mix, and ddH2O to a final volume of 50 µl, after which 1 µl of PfuUltra HF DNA polymerase (2.5 U/µl) was added. Oligonucleotide primers 1 and 2 for each mutation were as follows:

	p.Arg156Leu	p.GLu161Lys
FOR	5'-GGCTCAACTTTCTC**CTC**CTGGTGCACTTGGA-3'	5'-CGCCTGGTGCACTTG**AAA**GGAAATCAGCTCA-3'
REV	5'-TCCAAGTGCACCAG**GAG**GAGAAAGTTGAGCC-3'	5'-TGAGCTGATTTCC**TTT**CAAGTGCACCAGGCG-3'

The reaction mixes were cycled under the following conditions: after an initial denaturation step of 30 sec at 95°C, cycling was performed (for 18 cycles for control and 12 cycles for reaction) at 95°C for 30 sec, 55°C for 1 min and 68°C for 1 min/kb plasmid length. Following temperature cycling, the reaction was placed on ice for 2 minutes to cool to ≤37°C. 1 µl of the Dpn I restriction enzyme (10 U/µl) was then added to each amplification reaction and incubated at 37°C for 1 hour to digest the parental (non-mutated) supercoiled dsDNA. 1 µl of the Dpn I-treated DNA from each control and sample reaction was then transferred to 50 µl

of XL1-Blue supercompetent cells and incubated on ice for 30 min, followed by heat shock for 45 sec at 42°C and 2 min on ice. 0.5 ml of LB broth preheated to 42°C was then added and the transformation reactions incubated at 37°C for 1 hour with shaking at 225–250 rpm. Cells were then plated on LB-ampicillin agar plates and incubated overnight at 37°C. Single colonies were picked and grown in LB medium-ampicillin before plasmid extraction (QIAprep spin Mini-prep kit used according to the manufacturer's instructions, Qiagen, Crawley, UK, see Appendix 5 for reagents). Mutations were verified by sequencing.

HEK293 (sourced from ATCC, Middlesex, UK) were transfected with wild type or mutant plasmids using Lipofectamine-2000 reagent (Invitrogen, Paisley, UK), see section 2.5.6. Transfected cells were selected with G418 (1mg/ml, Sigma Aldrich, Poole, UK). Once stable cell lines were established, cells were cultured for 24 hours in 6 well plates to full confluency in 1ml of media. This process was repeated on 3 separate occasions with cells from the same stable cell lines, to produce technical replicates. Conditioned media from each well was removed for analysis before cell lysis as described in section 2.6.1. Lysates and conditioned media were cleared by centrifugation at 13000 rpm, 4°C, 10 min. Equal amounts of lysates and conditioned media were size-separated (NuPage BisTris gels 4-12%) and transferred to nitrocellulose membranes (Promega, Southampton, UK) as described in section 2.6. Post transfer the blots were stained with Ponceau red to assess equal loading, blocked with PBS containing 0.1% Tween-20 and 5% non-fat dry milk and incubated in a 1:500 dilution of mouse monoclonal anti-GFP (Roche Diagnostics Ltd, West Sussex, UK) antibody or a 1:3000 dilution of mouse monoclonal anti-GAPDH (Santa Cruz Biotechnology, Heidelberg, Germany) for 12 hours at 4°C. After washes, membranes were incubated with goat anti mouse IRDye680

(1:10000 dilution; Licor). Immunoblots were scanned with, and protein relative amounts were calculated by, the Odyssey ® Fc Imaging System (Licor).

2.10.4 Non-radioactive in situ hybridization (NR-ISH)/ Immunohistochemistry

E10.5 to E17.5 mouse embryos were collected from timed-crosses of wild type C57BL/6 mice. The morning of the vaginal plug was designated 0.5 days. 11 post conceptual weeks (pcw) human brains were obtained from the MRC-Wellcome Human Development Biology Resources (HDBR- Institute of Genetics Medicine, Newcastle, UK). Tissues were fixed in 4% paraformaldehyde (PFA) in PBS, cryoprotected in 30% sucrose, and frozen in Optimal Cutting Temperature (OCT) compound (VWR, Lutterworth, UK). 12-µm thick serial sagittal and coronal sections were cut using a cryostat and collected on Superfrost Plus slides (VWR, Lutterworth, UK).

Mouse and human *Igsf10* were PCR amplified from brain cDNAs using the following primers:

	Mouse-Igsf10	Human-IGSF10
Forward	5'-GCAAGAAGGAAAGAATCCCC -3'	5'-TCAGGAGCTTGACACGATTG-3'
Reverse	5'- GATTCGCCCATCCTCACTAA -3'	5'- CTGCGGTGTTTCACTAAGCA-3'

Amplified cDNAs were cloned into the dual promoter vector pGEM-T easy (Promega, Southampton, UK) and linearized with the appropriate restriction enzymes. Probe preparation was as follows. Digoxigenin (DIG)-labelled antisense and sense cRNA probes were synthesized by *in vitro* transcription in the presence of DIG-labelling mix (Roche Diagnostics Ltd, West Sussex, UK) using ~1 µg of linearized template and T7 or SP6 RNA polymerase (New England Biolabs). The concentration and integrity of each RNA probe was

analysed by gel electrophoresis and spectrophotometrically; for each probe, the transcription reaction was diluted with diethylpyrocarbonate (DEPC)-treated H_2O to a concentration of 100 ng/µl of transcript, aliquoted, and stored at −80°C. All probes were used at a concentration of 400–800 ng/ml of hybridization buffer. Each probe was designed to include regions of low nucleotide identity with other related family members or other sequences located in the NCBI nucleotide database. Mouse Cxcl12 probe was from Memi *et al* (247), as a gift from Dr Anna Cariboni.

The *in situ* protocol was performed as in Guasti *et al* (248). Sections were fixed in PFA, washed in PBS-Tween 0.1% (T-PBS), and permeabilised with proteinase K (Sigma Aldrich, Poole, UK, 5ug/ml in 100mM Tris HCL pH 7.5 and 50mM EDTA, pH 8.0) for 10 min at 37 °C. After re-fixing in PFA and 3 washes, sections were hybridized overnight at 58°C with either sense or antisense riboprobes in incubation buffer containing 50% formamide, 0.3M NaCl, 10mM Tris-HCL pH7.5, 1mM EDTA, 5% dextran sulphate, 1 x Denhardt's solution, 0.5mg/ml denatured salmon sperm DNA and 0.02% SDS. Slides were then washed through a series of saline-sodium citrate (SSC, National Diagnostic) buffers (2x SSC, three times each for 30 min, and once each in 1x SSC, 0.2x SSC and 0.05x SSC for 15 min, respectively) at 65°C. They were then washed in STE buffer (0.5M NaCl, 10mM Tris pH 7.5, 5mM EDTA) for 10 min at RT followed by treatment with RNase A (QIAGEN, 25ml of 20mg/ml stock in 50 ml of STE buffer) for 30 minutes at 37°C. After washing the slides twice with Maleic Acid Buffer (MAB) (0.1M maleic acid, 0.15 NaCl, pH 7.5, 0.1% Tween) they were incubated with anti-DIG alkaline phosphatase-Fab fragments (Roche Diagnostics Ltd, West Sussex, UK) diluted 1:2000 with 0.5% Blocking Reagent (Roche Diagnostics Ltd, West Sussex, UK) in MAB/0.1% Tween 20, overnight at 4°C.

After two washes with T-PBS the slides were equilibrated in alkaline buffer (100mM Tris pH 9.5, 100mM NaCl, 50mM $MgCl_2$, 1% Tween-20) for 10 min before being incubated at RT with NBT/BCIP (Roche Diagnostics Ltd, West Sussex, UK) in alkaline buffer supplemented with levimasole (Vector Laboratories, Peterborough, UK; 1drop/5ml). Sections were checked every hour until there was adequate staining. Sections were finally washed in T-PBS and mounted with mounting medium (glycerol/PBS, 3:1 v/v).

When co-labelling was desired after *in situ*, sections were incubated with primary antibodies (anti-GnRH, Immunostar, Wisconsin, USA; anti-peripherin, Merck-Millipore, Watford, UK; anti isolectin B4, Sigma-Aldrich, Poole, UK) diluted 1:1000 in PBS-Triton 0.1%, overnight at RT. After three washes with PBS-Triton 0.1%, slides were incubated for 2 hours at RT with biotin-conjugated goat secondary antibodies (Vector Laboratories, Peterborough, UK), diluted 1:300 in PBS and, after further washes, with the avidin-biotin complex (ABC staining kit, Vector Laboratories, Peterborough, UK). Sections were reacted with 4',6'-diamino-2-phenylindole (DAPI, Vector Laboratories, Peterborough, UK) and mounted as above.

Images were acquired using a Leica DM5500B microscope (Leica, Nussloch, Germany), equipped with a DCF295 camera (Leica) and DCViewer software (Leica), and then processed with Abode Photoshop CS6 and Adobe Illustrator CS6.

2.10.5 Migration experiments

Igsf10 silencing was achieved in NIH3T3 cells (sourced from ATCC, Middlesex, UK) using commercially produced puromycin-resistant synthetic 'SureSilencing' shRNA plasmids (Qiagen, Crawley, UK). Introduction of small double-stranded synthetic RNA can induce sequence-specific gene silencing in mammalian systems, and small hairpin 'shRNAs' are

used to achieve longer term gene-silencing. These are produced as single-stranded molecules of 50-70 nucleotides in length comprising a 5-10 nucleotide loop connecting 2 complementary 19-29 nucleotide long RNA fragments that create a double stranded stem by base pairing. These form a stem loop or 'hairpin' structure *in vivo*. In the cytoplasm the shRNA are converted, by cellular nucleases and the multidomain ribonuclease DICER, into small fragments (siRNAs). DICER then loads these small RNA products (siRNA) into a large multiprotein complex called the RNA-induced silencing complex (RISC) during which the passenger strand is cleaved by the enzyme Argonaute 2 (Ago2). The 78 antisense strand then guides the RISC to the corresponding mRNA as a result of sequence homology and the same Ago2 nuclease cleaves the target mRNA achieving sequence specific gene silencing. With continual expression of siRNAs toxic effects can occur. To account for these, controls (called scrambled shRNA) are generated using a shRNA insert that does not target human genes. Both *Igsf10*-specific and scrambled shRNA plasmids were transfected into NIH3T3 cells according to the manufacturer's instructions. The level of *Igsf10* knockdown was assessed by RT-qPCR (see section 2.12.2). Silencing was achieved at 80% in *Igsf10*-silenced NIH3T3 cells compared to cells expressing the scrambled shRNA. Stable cell lines were obtained with puromycin treatment (2 μg/ml).

Micromass cultures were obtained with the immortalized GnRH-expressing cell line GN11 (gift from Dr Anna Cariboni). Confluent monolayer cultures of GN11 were released by trypsin-EDTA, and re-suspended in growth medium. Micromasses were obtained by pipetting 20 μl (4.0×10^5 cells) of cell suspension onto the lid of a 10 cm petri dish and incubating for 48 hours in a 5% CO_2 incubator at 37°C. Micromasses were then placed onto confluent scrambled and *Igsf10*-silenced NIH3T3 cells for 7 days. Cells were then fixed in 4% PFA and GN11 migration into NIH3T3 was assessed by staining with anti-GnRH antibodies.

Migration index was calculated by assessing GnRH-stained cells using the ImageJ software (NIH). Statistical significance was evaluated by ANOVA and Student's t test. $P < 0.05$ (*) and $p < 0.01$ (**) were taken to be significant. Data are presented as mean ± standard error of mean and expressed as percentage reduction of migration, taking untreated or scrambled controls as 100%. Each experiment was performed using at least 12 micromasses/group and repeated three times.

2.10.6 Zebrafish Investigations

Animals: I performed initial zebrafish *in situ* hybridisation experiments on zebrafish embryos obtained from Zebrafish Neurobiology and Behaviour Group, School of Biological and Chemical Sciences, Queen Mary, University of London. All further zebrafish experiments were carried out at the Department of Neurobiology, Tel-Aviv University, by Dr Valentina Andre under the supervision of Professor Yoav Gothilf. A transgenic line, *Tg(gnrh3:EGFP)*, was used to visualize GnRH3 neurons and their projections (249). Wild type and transgenic zebrafish embryos were generated by natural mating. Both AB and TL strains were used for the experiments. Embryos were raised in fish water with methylene blue (0.3 p.p.m.) in Petri dishes at 28°C in a 12 hours light/12hours dark cycle. Developmental stages are expressed in hours post fertilization (hpf) or days post fertilization (dpf). All embryos used in experiments were 5 dpf or younger. Pigmentation was prevented by adding 0.2 mM phenylthiourea (PTU) to the water at 24 hpf.

RT-PCR analysis: Total RNA was isolated from embryos and larvae at different stages of embryogenesis using EZ RNA Total RNA Isolation kit (Biological Industries, Beit Haemek, Israel). First-strand cDNA was synthesized from 1mg of total RNA by qScript cDNA kit

(Quanta Biosciences, Gaithersburg, MD, USA) according to the manufacturer's protocol. Primers for Zebrafish *Igsf10* were as follows:

RT-PCR	Forward	Reverse
Zf-Igsf10	5'-TTGGCTACAGTCCCGATTTC-3'	5'- AAATTTTGCTGGGACGAATG-3'

In situ hybridisation experiments: Zebrafish *Igsf10* were PCR amplified from brain cDNAs using the following primers:

ISH	Forward	Reverse
Zf-Igsf10	5'-CTTCCCGACCACACAGTTTT-3'	5'-TCATTTATACCATCCCCCGA-3'

Amplified cDNAs were cloned into the dual promoter vector pGEM-T easy (Promega, Southampton, UK) and linearized with the appropriate restriction enzymes. Probe preparation was performed as described in section 2.10.4. Dechorionated embryos were fixed in 4% PFA/PBS overnight at 4°C and then washed in 100% MeOH and stored at -20°C. For rehydration, embryos were washed in decreasing concentrations of MeOH/DEPC PBST and then underwent 5 x 5 min washes with DEPC PBST. Permeabilisation with proteinase K (stock 2mg/ml) was titrated for the age of embryos (<24 hrs – no PK, 24hr – 10ug/ml for 20mins at RT, 36hr – 20ug/ml for 25mins at RT, 48hr – 25ug/ml for 25mins at RT, 60hr – 25ug/ml for 30mins at RT, 72hr – 50ug/ml for 25mins at RT, 96hrs 100ug/ml for 30mins at RT). Embryos were then washed with DEPC PBST for 5min x 2 times at RT and refixed in 4%

PFA/PBS at RT. Repeat washes with DEPC PBST at RT were followed by incubation with hybridisation buffer for 2hrs at 65°C (Hybridisation buffer: 50% formamide, 0.3M NaCl, 10mM Tris-HCL pH7.5, 1mM EDTA, 5% dextran sulphate, 1 x Denhardt's solution, 0.5mg/ml denatured salmon sperm DNA and 0.02% SDS). Probes were heated in hybridisation buffer for 5mins at 68°C and chilled on ice. Incubation solution was replaced with 300ul hybridisation buffer containing 150ng RNA probe and embryos incubated at 65°C overnight. Post-hybridisation washes were then carried out at 65°C (10min each) with decreasing concentrations of hybridisation buffer/2xSSC. Following 4 further PBST washes at RT the embryos were incubated with STE buffer for 10 mins, followed by RNAse treatment for 30mins at 37°C (RNase A 25ul of 20mg/ml stock in 50ml of STE buffer). Embryos were then incubated in PBST containing 2% sheep serum and 2mg/ml BSA with gentle agitation for 1hr at RT, followed by overnight incubation at 4°C with preabsorbed antiDIG-AP antibody in PBT containing 2% sheep serum and 2mg/ml BSA (1/5000 dilution). 3 x 5min washes with BCL buffer (0.1M Tris/HCL pH 9.5, 0.1M NaCl, 0.1% Tween-20) was followed by incubation (in dark) with BCL buffer including 4.5ul NBT, 3.5ul BCIP per lm. Post staining the embryos were washed and refixed in PFA/PBS for 2hrs.

Morpholino knockdown experiments: Morpholino-modified antisense oligonucleotides (MOs) (Gene Tools, Philomath, OR, USA) were used to repress *Igsf10* mRNA to assess the effect on the development of the GnRH3 system. MOs were selected for knockdown of our candidate gene because of the ease of delivery and their high efficacy throughout zebrafish embryonic and larval development. MOs are a synthetic derivative of DNA with two major changes: (i) a six-membered morpholine ring replaces the standard deoxyribose ring, and (ii) a non-ionic phosphorodiamidate linkage replaces the anionic phosphodiester bond (229). As they are neutrally charged and of relatively small size (usually 25 bases in length), diffusion is the main

mechanism of spread throughout the embryo following microinjection. MOs are extremely effective through 50 hpf but can lose efficacy in a sequence specific manner thereafter.

MO activities in zebrafish embryos include both sequence-specific RNA binding as well as effects not associated with loss of function of the targeted locus. The latter 'off-target' effects can be a confounding variable when using MOs for the assignment of function to sequence in zebrafish. Thus, control MOs are used in parallel to allow assessment of these off-target effects. A splice site-MO (Igsf10 Sp-MO) and an ATG-MO (Igsf10 ATG-MO) were designed and ordered from Gene Tools LLC (Oregon, USA) with the following sequences:

	Zf-Igsf10 Morphino sequences
Splice Site	5'-GCCTGTTGTAGGTTTTACCCCAGGT-3', 25 bp
ATG (Start site)	5'-GGGAATCCGCTGCTGGGTCACACAT-3', 25 bp

Quantities of 1.5 pmol/embryo (Igsf10 Sp-MO) or 1 pmol/embryo (Igsf10 ATG-MO) were injected into *Tg(gnrh3:EGFP)* embryos immediately after fertilization. In all experiments, Igsf10 Sp-MO or ATG-MO injected embryos were compared with embryos injected with the same amount of mispair control (Control Sp-MO or Control (Ctrl) ATG-MO, both 25 bp, Gene Tools LLC, Oregon, USA) at the same developmental stage.

	Control Morphino sequences
Control Sp-MO	5'-GCGTCTTGTACGTTTTACCCGACGT-3', 25 bp
Control ATG-MO	5'-GAGAATACGCTACTGGGTAACAAAT-3', 25 bp

Two sets of experiments (biological replicates) were performed and each experiment consisted of 3 groups: Sp-MO injected, Ctrl-MO injected and uninjected. No statistical method was used to predetermine sample size. Rather, sample size was based on preliminary data and observed effect sizes. Dead embryos or embryos that were GFP negative at 48 hpf were excluded from the analysis. Considering that MO injection can cause a death rate of 10-25% in 24 hour old embryos, we injected a large number of embryos in order to establish statistical significance. In the first experiment the mortality rate in the injected embryos, Sp-MO and Ctrl-MO, was 11% and in the second was 25%. The mortality rate in the uninjected embryos was 2.5% and 5.7%, respectively.

In the first experiment 353 embryos were injected with Sp-Mo, 260 embryos were injected with Ctrl-MO and 244 embryos were left uninjected. 146/353, 108/260 and 100/244 embryos were available and GFP-positive for analysis at 48 hpf. In the second experiment 186 embryos were injected with Sp-Mo, 157 embryos were injected with Ctrl-MO and 122 embryos were left uninjected. 55/186, 52/157 and 56/122 embryos were available and GFP-positive for analysis at 48 hpf. Sex of embryos was undetermined (separate sexes can be detected only after 21-23 dpf).

Injected larvae were checked daily and the EGFP expression was monitored under SZX12 fluorescent dissecting microscope (Olympus, Tokyo, Japan) for 5 days. Quantification of the percentage of embryos showing an abnormal GnRH3-neuron phenotype in Sp-MO injected

embryos compared to relative controls, injected either with Ctrl-MO or uninjected was assessed by mean ± standard error of mean, tested by one-way ANOVA followed by Dunnett's post-hoc test.

Efficacy of Igsf10 Sp-MO was evaluated by RT-PCR. Injected and uninjected embryos were sampled at 48 hpf and total RNA was extracted (EZ RNA Total RNA Isolation kit, Biological Industries, Beit Haemek, Israel). 1mg of RNA was retro-transcribed using qScript cDNA kit (Quanta Biosciences, Gaithersburg, MD, USA). Specific primers corresponding to zebrafish *Igsf10* exons 2 and 3 (Igsf10-E2 FOR and Igsf10-E3 REV) were used for PCR amplification:

	Zf-Igsf10
Forward	5'-GCGGATTCCCGTCTGCTATGG-3'
Reverse	5'-TGCAGAGATGTGAGGCCACTGAAC-3'

PCR products were separated on an agarose gel, extracted from gel using HiYield Gel/PCR DNA Fragments Extraction Kit (RBCBioscience, New Taipei City, Taiwan) and sequenced. Sequencing analysis demonstrated the insertion of 97 nucleotides within the intron 2 of Igsf10 Sp-MO-injected embryos, which would result in a frameshift and produce a dysfunctional Igsf10 protein in these embryos. (Fig. 2.10.6).

Figure 2.10.6 - Efficacy of Igsf10 Sp-MO. Nucleotide sequence obtained from RNA of Igsf10 Sp-MO injected embryos showing an insertion of 97 nucleotides in the splice site of the second exon. Apostrophes denote homology.

2.11 Functional Annotation of *HS6ST1*

2.11.1 Assessment of tissue expression of HS6ST1

Cells and tissue expression: RNA extraction (RNeasy Mini kit, QIAGEN) and cDNA synthesis (M-MLV Reverse Transcriptase, Promega, Southampton, UK) were performed according to the manufacturers' instructions (see Appendix 5 and section 2.4 for detailed methods) using adult mouse hypothalamus tissue (whole). Reverse transcription PCR was performed on diluted cDNA samples with Sigma Taq DNA Polymerase using the following program: 95 °C for 5 minutes, 40 cycles at 95 °C for 15 seconds, 60 °C for 30 seconds, and 72 °C for 30 seconds. Primers used were:

	mHs6st1
Forward	5'-ACTGGACCGAACTCACCAAC-'3
Reverse	5'-GCAGCAGGGTGATGTAGTA-'3

Non-radioactive in situ hybridization (NR-ISH)/ Immunohistochemistry: E11-17.5 mouse embryo and 9 post conceptual weeks (pcw) human brain sagittal and coronal sections were collected as described in section 2.10.4. Mouse and human *Hs6st1* were PCR amplified from brain cDNAs using the following primers:

	mHs6st1	hHS6ST1
Forward	5'-ACTGGACCGAACTCACCAAC-3'	5'-TGATCGTCTTCCTGCACATC -3'
Reverse	5'-AACTCAGTGAGGCCGAAGAA-3'	5'- CGCGTGCTATTGTACTGCAT -3'

Amplified cDNAs were cloned into the dual promoter vector pGEM-T easy (Promega, Southampton, UK) and linearized with the appropriate restriction enzymes. Probe preparation and *in situ* protocol was as in section 2.10.4.

When co-labelling was desired after *in situ*, the sections were incubated with primary antibodies (anti-GnRH, Immunostar, Wisconsin, USA; anti-peripherin, Merck-Millipore, Watford, UK) diluted 1:1000 in PBS-Triton 0.1%, overnight at RT. After three washes with PBS-Triton 0.1%, slides were incubated for 2 hours at RT with biotin (Vector Laboratories, Peterborough, UK) or Alexa 568-conjugated (R&D Systems, UK) goat secondary antibodies, diluted 1:300 or 1:500 respectively in PBS. Biotin-conjugated antibodies were washed with avidin-biotin complex (ABC staining kit, Vector Laboratories, Peterborough, UK) and the sections developed with 4',6'-diamino-2-phenylindole (DAPI, Vector Laboratories, Peterborough, UK) and mounted as above. Sections stained with Alexa568-conjugated

antibodies were mounted directly in Dako Mounting Medium (Dako). Images were acquired and processed as described in section 2.10.4.

2.11.3 *In vitro* assay of sulphotransferase activity

The human HS6ST1 cDNA (IMAGE consortium accession ID BC099638) was obtained from Genecopoeia in a pReceiverM14a backbone (pReceiverM14a-WThHS6ST1). The p.Arg375His and p.Met404Val point mutations were introduced into pReceiverM14a-WThHS6ST1 by PCR-mediated mutagenesis (QuikChange II Site-Directed Mutagenesis Kit, Agilent Technologies, Stockport, UK, see Appendix 5 and section 2.10.3) using the following primers:

	p.Arg375His	p.Met404Val
Forward	5'-GCCTGAGGAGC**CAC**GAGGAGCGTCT-3'	5'-CCCACCGAGGACTAC**GTG**AGCCACATCATTG-3'
Reverse	5'-AGACGCTCCTC**GTG**GCTCCTCAGGC-3'	5'-CAATGATGTGGCT**CAC**GTAGTCCTCGGTGGG-3'

They were verified by sequencing. 90% confluent COS7 cells in 6 well plates were transfected with 2 μg of pReceiverM14a-WThHS6ST1 or pReceiverM14a-R375H-MThHS6ST1 or pReceiverM14a-M404V-MThHS6ST1 and 0.5 μg of pcDNA4TO-eGFP (as a transfection control) using Lipofectamine 2000 (Invitrogen, Paisley, UK, see section 2.5.6). Transfected cells were cultured for 48h before lysis on ice using chilled lysis buffer (TBS 1% Triton X100 - 50 mM Tris HCl, pH 7.4, with 150 mM NaCl, 1 mM EDTA, and 1% Triton X-100). The homogenates were stirred for 1 hr and then centrifuged at 10,000 × g for 30 seconds at 4°C. HS6ST1-FLAG fusion proteins were isolated from the supernatant (cell extract) by anti-FLAG M2 affinity chromatography (Sigma Aldrich, Poole, UK) according to the manufacturer's instructions. In brief, the sample was passed through an anti-FLAG M2 affinity column,

washed and then eluted using 0.1M glycine HCL into 1M TRIS. The purity of the eluted protein was checked by Coomasie brilliant blue G-250(5x) (Bio-Rad, Hemel Hempstead, UK) staining of SDS-page electrophoretic gels.

Protein relative amounts were calculated by densitometric analysis of Western blots. Western blotting was as described in section 2.8. All incubations were performed for 1 hr at RT in blocking buffer. The membranes were incubated in a 1:1000 dilution of mouse monoclonal M1 anti-FLAG antibody (Sigma Aldrich, Poole, UK). Secondary horseradish peroxidase-conjugated mouse (Cell Signalling Technology, Hitchin, UK) antibody was used at a concentration of 1:10,000.

Determination of sulphotransferase activity: Sulphotransferase activity was measured using a universal sulphotransferase activity colorimetric assay (R&D systems, UK, see Appendix 5) according to the manufacturer's instructions. Briefly, different amounts of purified HS6ST1 were added to a mix of 0.2 mM 3'-Phosphoadenosine-5'-phosphosulfate (PAPS, donor), 1 mM N-acetylglucosamine (NAcGluc, acceptor) and 10 ng/ml inositol monophosphatase 3 (coupling enzyme) in a total volume of 50 ml of reaction buffer. The elution product of mock-transfected cells was used as negative control. The reaction was allowed to proceed for 20 min at 37°C and inorganic phosphate (Pi) was measured immediately after by the Malachite Green Method using reagents provided in the universal sulphotransferase activity kit. Enzymatic activity was represented as percent of wild type activity (after subtracting the blank), normalised by the amount of protein. Statistical significance was evaluated by 2-tailed student t-test, with $p<0.05$ (*) taken to be significant. Each experiment had 3 biological replicates.

2.12 Functional Annotation of *FTO*

2.12.1 Assessment of Kinetic Activity of mutant FTO proteins

In silico analysis: The FTO experimentally solved structure (PDB identifier: 4cxx) was used to study the structural effect of FTO variants. The following interactions involved in protein stability were considered: i) salt bridges, defined as at least one pair of atoms on oppositely charge groups within a 4.5 Å distance; ii) hydrogen bonds (H-bond), defined as a donor–acceptor distance ≤2.5 Å and an angle at the acceptor ≥90°; and iii) disulphide bridges (S-S bridge) defined as the side chains of two cysteines at a 3.0 Å distance. N-glycosylation sites were determined based on the consensus sequence Asn-X-Thr/Ser (X= any amino acid, except proline). The DSSP program was used to calculate surface accessibility and Disopred3 (250) to predict disordered protein regions.

In vitro studies of FTO mutant proteins: Cloning of wild-type human FTO cDNA into pET302/NT-His (Invitrogen, Paisley, UK) has been described previously (251), and the wild-type plasmids were sent to us as a gift by Dr Giles Yeo. The p.Leu44Val and p.Ala163Thr point mutations were introduced using PCR-mediated mutagenesis (QuikChange II Site-Directed Mutagenesis Kit, Agilent Technologies, Stockport, UK, see Appendix 5 and section 2.10.3) using the following primers:

	p.Leu44Val
Forward	5'-GAATTCTATCAGCAGTGGCAG**GTG**AAATATCCTAAACTAATTCT-3'
Reverse	5'-AGAATTAGTTTAGGATATTT**CAC**CTGCCACTGCTGATAGAATTC-3'

	p.Ala163Thr
Forward	5'-TTGCTGCCAAAGAGAAG**ACT**AATGAGGATGCTGTG-3'
Reverse	5'-CACAGCATCCTCATT**AGT**CTTCTCTTTGGCAGCAA-3'

These were verified by sequencing. Recombinant wild-type and mutant FTO expression plasmids were transformed into *Escherichia coli* BL21-Gold (DE3) (Stratagene) and cultured in LB broth and 50 μg/ml carbenicillin. Expression of the cloned gene was induced by the addition of 0.5 mM IPTG (Thermofisher Scientific) at 15°C for 4 hrs. The cells were harvested and pellets resuspended in lysis buffer (50 mM Hepes/KOH (pH 8.0), 2 mM 2-mercaptoethanol, 5% glycerol and 300mM NaCl] before digestion with lysozyme (1 mg/ml). The cleared lysate was supplemented with imidazole (final concentration 10 mM) before mixing with 1 ml of pre- washed Ni-NTA (Ni^{2+}-nitrilotriacetate) beads (Qiagen, Crawley, UK). After binding for 1 hr, the mixture was re-loaded on to the column and washed with lysis buffer supplemented with increasing concentrations of imidazole. FTO was eluted with 2 ml of lysis buffer containing 250 mM imidazole. The eluate was concentrated with a 30 kDa MWCO (molecular-mass cut-off) concentrator (Sartorius Stedim) with buffer changing to 20 mM Hepes/KOH (pH 8), 5 % glycerol and 50 mM NaCl. Purified proteins were snap-frozen and stored at −80°C. Protein purity was checked by Coomasie brilliant blue staining of SDS-page electrophoretic gels.

Fluorescent FTO 3-meU demethylation assay: An RNase-cleavage assay, previously verified by Dr Giles Yeo's group (252), was used to measure the demethylation activity of the 2 FTO mutant proteins on 3-meU. This experiment was carried out by Dr Marcella Ma, MRC Metabolic Diseases Unit, Cambridge. After initial optimization testing, in each reaction, 100

nM of either 14C-labelled 3-methylthymidine- labelled DNA substrate or unmethylated substrate was mixed with 75 µM Fe(NH4)2(SO4)2, 300 µM 2-OG, 2 mM ascorbate, 50 µg/ml BSA and various amount of RNase A from 10 − 12 µg/µl to 0.1 µg/µl in 50 mM Tris/HCl buffer at pH 7.0. Samples were prepared in duplicate and the FAM (6-carboxyfluorescein) emission was measured for 30 min at a wavelength of either 520 nm or 535 nm (depending on the plate reader) with excitation at 485 nm. The measurement was performed using a microplate reader [BMG Pherastar (520 nm), Labtech, or Infinite f500 (535 nm), Tecan] in a dark flat-bottomed 96-well plate at RT (25°C).

2.12.2 Assessment of Vaginal Opening in FTO$^{+/-}$ and wild type mice

All FTO mice experiments were carried out by Dr Antony Coll and Dr Deb Rimmington at the MRC Metabolic Diseases Unit, Cambridge. For the vaginal opening study female *Fto* heterozygous mice (*Fto*$^{+/-}$) and their WT littermates were taken predominantly from either male *Fto*$^{+/+}$ x female *Fto*$^{+/-}$ or male *Fto*$^{+/-}$ x female $^{+/+}$ breeding pairs. Two litters of WT females (n= 5 and n= 2 respectively) with no heterozygous littermates were also included in the study. From P21, the day of weaning, all female mice were weighed and visual examination of the vagina was carried out by placing the mouse on top of a cage lid and lifting the tail vertically away from the body. No excessive force was involved. Weighing and visual examination continued until 1 day after 1st opening.

For all experiments, data are expressed as the mean ± SEM. To determine statistical significance, we used the unpaired t test (2-tailed). A p value of less than 0.05 was considered statistically significant. Statistical analysis was performed using SPSS Software.

2.13 *In vitro* analysis of pathogenicity of variants in *IGSF1* identified in DP patients

These experiments were carried out in the lab of Professor Wit, Leiden University Medical Center, by members of their group. Briefly, HEK293 cells were transfected with expression vectors specific for each *IGSF1* variant, followed by cell surface biotinylation and immunofluorescence to determine plasma membrane expression of the resulting IGSF1 proteins, as previously described (253).

2.14 Microarray analysis of rat neuronal tissue

Microarray analysis was designed and supervised by Dr Anna Cariboni and her group, and carried out at the University of Milan using a pipeline developed by the Genopolis consortium (www.genopolis.it). Analysis was performed on both GnRH:GFP primary rat neurons at E14, E17 and E20 after cell sorting, and GN11 and GT1-7 cells. The Affymetrix ragene10 data chip (ragene10sttranscriptcluster), assembled using data from public repositories, was used for annotation. Data handling was done using Bioconductor packages. The Robust Multichip Average (RMA) method was employed to calculate probe set intensity. To filter out noisy data before the selection of differentially expressed genes a filter was applied based on an Inter Quantile Range (IQR), selected as greater than 0.01. The identification of differentially expressed genes was addressed using Fold Change. This method uses the logarithm of the ratio between each average of experimental conditions and the average of the condition chosen as the baseline (logRatios). For each experimental condition, the probesets whose fold change were greater than or equal to 2 were selected as significant.

Chapter 3

Results 1: Mutations in *IGSF10* Cause Self-Limited Delayed Puberty, via effects on GnRH neuronal migration

3.1 Rare, potentially pathogenic variants in the *IGSF10* gene found in 10 families with DP.

Initial whole exome sequencing performed in the 18 most extensive families from our cohort (111 individuals: a total of 76 individuals with DP, male=53, female=23; and 35 controls, male=13, female=22), identified 2,474,145 variants after quality control (Fig. 3.1.1). Following filtering through our in-house pipeline to identify rare, predicted deleterious mutations, segregating with trait in an autosomal dominant inheritance pattern in multiple families and with potential biological relevance, 28 top candidate genes were identified. These 28 genes were then put forward for targeted re-sequencing in a further 42 families from the same cohort (178 individuals with DP and 110 controls, Fig. 3.1.1), and the filtered results analysed by applying statistical thresholds for enrichment of rare, pathogenic variants in our cohort via whole gene rare variant burden testing with multiple comparison adjustment (238) (Table 3.1.1).

Four genes initially passed the adjusted $p<0.025$ threshold after rare variant burden testing, and potentially pathogenic variants in these genes were further analysed to determine their presence in controls from our cohort and for segregation with trait (Table 3.1.1 and Fig. 3.1.1).

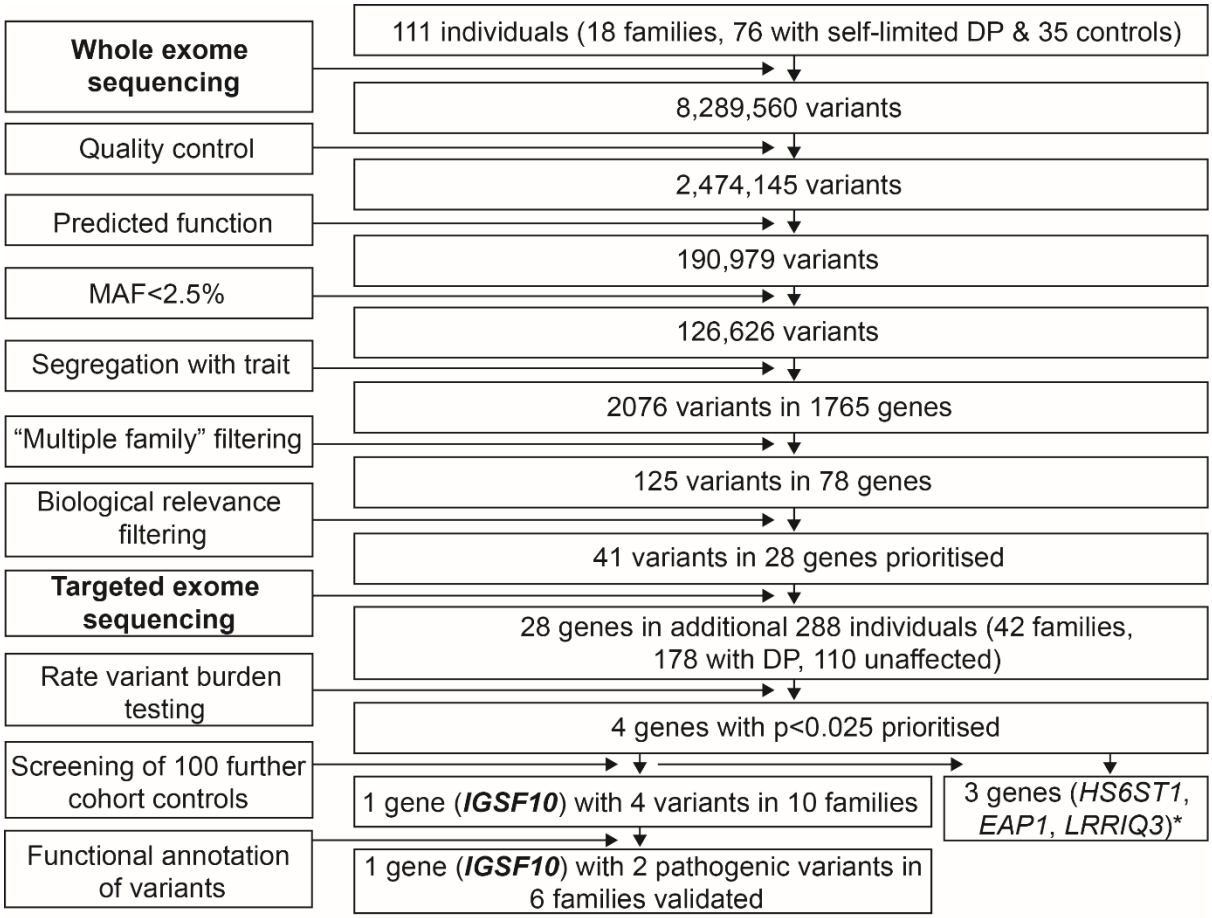


Figure 3.1.1 - Flowchart of exome sequencing filtering outcomes.

Whole exome sequencing was initially performed on DNA extracted from peripheral blood leukocytes of 111 individuals from 18 families from our cohort (76 with DP and 35 controls). The exome sequences were aligned to the UCSC hg19 reference genome. Picard tools and the genome analysis toolkit were used to mark PCR duplicates, realign around indels, recalibrate quality scores and call variants. Variants were analysed further and filtered for potential causal variants using filters for quality control, predicted functional annotation, minor allele frequency (MAF), segregation with trait, variants in multiple families and biological relevance (See methods and Table 3.1.1 for further information on filtering criteria). Targeted exome sequencing using a Fluidigm array of 28 candidate genes identified post-filtering was then performed in a further 42 families from the same cohort (288 individuals, 178 with DP and 110 controls). Variants post targeted re-sequencing were filtered using the same criteria as the whole exome sequencing data. Rare variant burden testing was performed for all genes selected for targeted re-sequencing, in order to rank candidate genes post targeted re-sequencing. A multiple comparison adjustment was applied to the set of 28 p values post hoc (238). Screening of 100 further cohort controls was via conventional Sanger sequencing. Functional annotation of the variants as described elsewhere in methods. DP – delayed puberty. *data shown elsewhere.

Gene	Total number of rare (MAF <2.5%) and predicted damaging* variant alleles in probands and in the reference population		P value	Adjusted P value**
	In DP cohort (n=49)	In ExAC Finn database (n=3305)		
HS6ST1	4	3	3.23E-06	3.01E-05
LRRIQ3	12	88	1.66E-07	4.65E-06
EAP1	8	39	1.42E-06	1.99E-05
IGSF10	13	292	0.0029	0.020
SEC24A	3	17	0.0051	0.029
ZNF560	2	6	0.00823	0.038
FTO	3	29	0.0188	0.058
LGR4	8	172	0.0155	0.058
CYFIP1	2	10	0.01852	0.058
TTF1	7	96	0.023	0.064
THBS4	3	33	0.02574	0.066
INHA	3	39	0.0384	0.090
SYPL2	5	114	0.04472	0.096
ELFN2	2	22	0.06745	0.135
SETBP1	2	25	0.0829	0.155
TBX6	4	102	0.1204	0.211
FANCL	3	78	0.175	0.288
ZNF266	2	54	0.236	0.367
HTR3D	2	60	0.3035	0.433
ICAM3	1	20	0.315	0.433
BCLAF1	4	393	0.3249	0.433
ARHGEF16	3	100	0.4329	0.551
CUX1	1	39	0.5141	0.626
CPZ	2	196	0.5867	0.684

PANX3	2	120	1	1
WDR25	2	126	1	1
ERCC4	3	202	1	1
LRRN1	0	4	1	1

Table 3.1.1: Rare Variant Burden Testing post Targeted Exome Sequencing

Minor allele frequency (MAF) data was retrieved from the ExAC Browser (Exome Aggregation Consortium (ExAC), Cambridge, MA: http://exac.broadinstitute.org, accessed October 2015). *by both SIFT (237) and Polyphen2 (236). P value calculated by Fisher's exact test. **Benjamini & Hochberg method (238)

Following this analysis, *Immunoglobulin superfamily member 10*, *IGSF10* (ENSG00000152580, gene identification number 285313), was found to be the most promising candidate to take forward for further analysis. Although the adjusted p value of 0.020 after rare variant burden testing was not the most significant of all the candidates tested, after screening of a further 100 controls from our cohort 4 potentially pathogenic variants in probands from 10 separate pedigrees were identified. The other 9 of 13 rare and potentially pathogenic variants identified in *IGSF10* after targeted exome sequencing were identified post-analysis as present in multiple sequenced controls from our cohort.

Variants in *LRRIQ3* were found in multiple sequenced controls from our cohort and were thus discounted. Analysis of variants identified in *HS6ST1* is discussed further in chapter 4. Analysis of variants in *EAP1* is currently in progress by another member of our lab group, Alessandra Mancini, as part of her PhD project.

Four variants in *IGSF10*, identified in 31 individuals from 10 families (NM_178822.4:c.467G>T (rs138756085) p.Arg156Leu, NM_178822.4: c.481G>A

(rs114161831) p.Glu161Lys, NM_178822.4: c.6791A>G p.Glu2264Gly and NM_178822.4: c.7840G>A (rs112889898) p.Asp2614Asn), were found in ≤1 control subject (Table 3.1.2).

Nucleotide change	Amino acid change	Exon	MAF from DP patients (%) (n=215)	MAF from controls (%) (n=210)	MAF (%) Finnish/European /All
c.467G>T	p.Arg156Leu	3	2.8	0	0/0.5/0.4
c.481G>A	p.Glu161Lys	3	5.6	0.5	2.0/0.7/1.0
c.6791A>G	p.Glu2264Gly	6	0.5	0	not seen
c.7840G>A	p.Asp2614Asn	6	3.3	0	0/0.8/0.8

Table 3.1.2 – Minor allele frequency of *IGSF10* variants in study population and control cohorts.

Minor allele frequency (MAF) data for the Finnish population was retrieved from The Sequencing Initiative Suomi (The SISu project) (http://www.sisuproject.fi/, release 3.0, accessed September 2015). European and other MAF data was retrieved from the ExAC Browser (Exome Aggregation Consortium (ExAC), Cambridge, MA: http://exac.broadinstitute.org, accessed September 2015).

Although 3 of the 4 variants were present in public databases, they were highly enriched in our cohort (Table 3.1.2). Analysis of self-limited DP families is complicated by the fact that this phenotype represents the tail of a normally distributed trait within the population, so it is anticipated that variants that govern the inheritance of this condition will also be present in the general population at a low level. Indeed, it is expected that up to 5% of the individuals sequenced in population databases will have abnormal pubertal timing, either early or delayed. Thus, the absence of these variants in population databases cannot be

used as an exclusion criterion and instead, a comparison of prevalence of such variants was made to identify those that are enriched in patients compared to the general population.

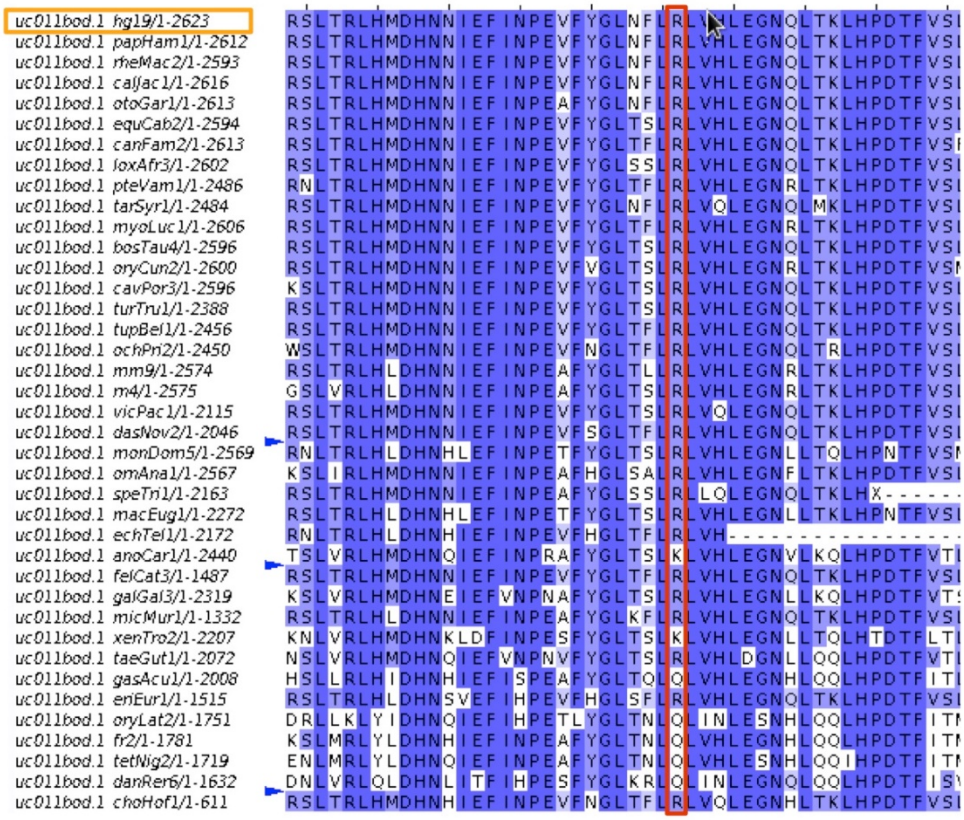
All four *IGSF10* variants are heterozygous missense variants predicted to be deleterious by ≥3/5 prediction tools (Table 3.1.3). All variants affect amino acids that are highly conserved among homologues, as revealed by PhyloP or GERP score, and multiple sequence alignment (Table 3.1.3 and Fig. 3.1.2).

AA Change	dbSNP137 ID	PhyloP (254)	SIFT (237)	PolyPhen 2 (236)	LRT (255)	Mutation Taster (256)	FATHMM (257)	GERP++ (258)
p.R156L	rs138756085	C	D	D	D	N	D	5.18
p.E161K	rs114161831	C	D	D	D	D	T	4.94
p.E2264G	n/a	C	D	P	N	D	T	3.71
p.D2614N	rs112889898	C	D	D	D	D	T	5.24

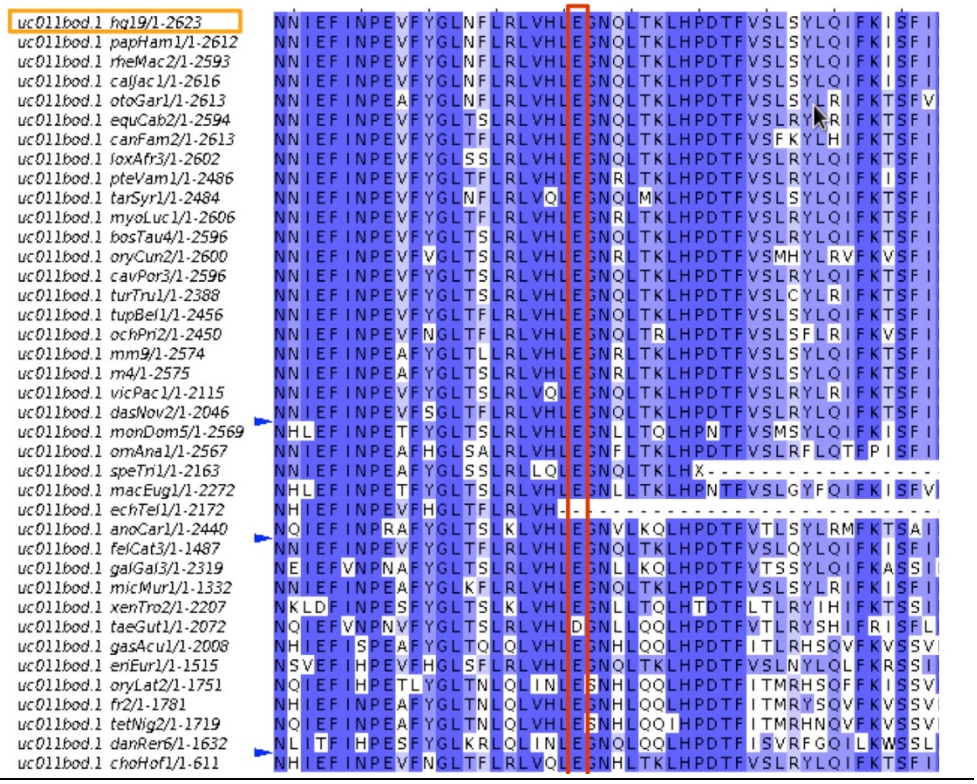
Table 3.1.3 – Prediction of *IGSF10* variants according to web-based prediction software programs and conservation across species.

C –conserved, D – deleterious, disease causing or damaging, P – possibly damaging, N – neutral, T – tolerated

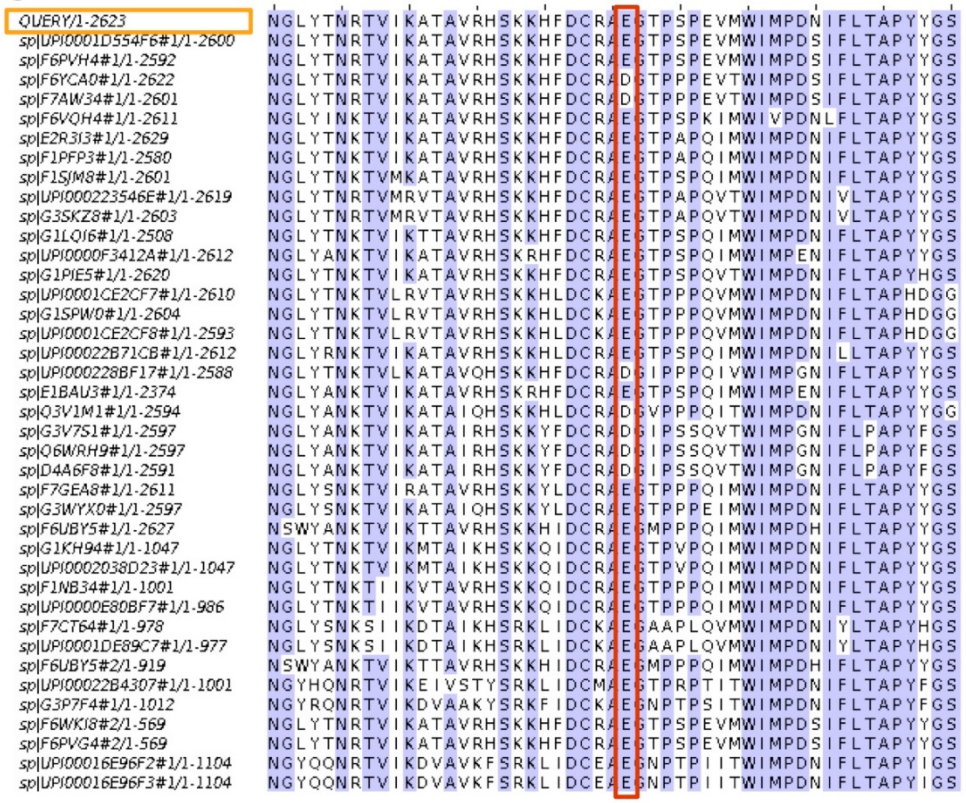
A


uc011bod.1 hg19/1-2623
uc011bod.1 papHam1/1-2612
uc011bod.1 rheMac2/1-2593
uc011bod.1 calJac1/1-2616
uc011bod.1 otoGar1/1-2613
uc011bod.1 equCab2/1-2594
uc011bod.1 canFam2/1-2613
uc011bod.1 loxAfr3/1-2602
uc011bod.1 pteVam1/1-2486
uc011bod.1 tarSyr1/1-2484
uc011bod.1 myoLuc1/1-2606
uc011bod.1 bosTau4/1-2596
uc011bod.1 oryCun2/1-2600
uc011bod.1 cavPor3/1-2596
uc011bod.1 turTru1/1-2388
uc011bod.1 tupBel1/1-2456
uc011bod.1 ochPri2/1-2450
uc011bod.1 mm9/1-2574
uc011bod.1 rn4/1-2575
uc011bod.1 vicPac1/1-2115
uc011bod.1 dasNov2/1-2046
uc011bod.1 monDom5/1-2569
uc011bod.1 omaAna1/1-2567
uc011bod.1 speTri1/1-2163
uc011bod.1 macEug1/1-2272
uc011bod.1 echTel1/1-2172
uc011bod.1 anoCar1/1-2440
uc011bod.1 felCat3/1-1487
uc011bod.1 galGal3/1-2319
uc011bod.1 micMur1/1-1332
uc011bod.1 xenTro2/1-2207
uc011bod.1 taeGut1/1-2072
uc011bod.1 gasAcu1/1-2008
uc011bod.1 eriEur1/1-1515
uc011bod.1 oryLat2/1-1751
uc011bod.1 fr2/1-1781
uc011bod.1 tetNig2/1-1719
uc011bod.1 danRer6/1-1632
uc011bod.1 choHof1/1-611

B


uc011bod.1 hg19/1-2623
uc011bod.1 papHam1/1-2612
uc011bod.1 rheMac2/1-2593
uc011bod.1 calJac1/1-2616
uc011bod.1 otoGar1/1-2613
uc011bod.1 equCab2/1-2594
uc011bod.1 canFam2/1-2613
uc011bod.1 loxAfr3/1-2602
uc011bod.1 pteVam1/1-2486
uc011bod.1 tarSyr1/1-2484
uc011bod.1 myoLuc1/1-2606
uc011bod.1 bosTau4/1-2596
uc011bod.1 oryCun2/1-2600
uc011bod.1 cavPor3/1-2596
uc011bod.1 turTru1/1-2388
uc011bod.1 tupBel1/1-2456
uc011bod.1 ochPri2/1-2450
uc011bod.1 mm9/1-2574
uc011bod.1 rn4/1-2575
uc011bod.1 vicPac1/1-2115
uc011bod.1 dasNov2/1-2046
uc011bod.1 monDom5/1-2569
uc011bod.1 omaAna1/1-2567
uc011bod.1 speTri1/1-2163
uc011bod.1 macEug1/1-2272
uc011bod.1 echTel1/1-2172
uc011bod.1 anoCar1/1-2440
uc011bod.1 felCat3/1-1487
uc011bod.1 galGal3/1-2319
uc011bod.1 micMur1/1-1332
uc011bod.1 xenTro2/1-2207
uc011bod.1 taeGut1/1-2072
uc011bod.1 gasAcu1/1-2008
uc011bod.1 eriEur1/1-1515
uc011bod.1 oryLat2/1-1751
uc011bod.1 fr2/1-1781
uc011bod.1 tetNig2/1-1719
uc011bod.1 danRer6/1-1632
uc011bod.1 choHof1/1-611

C


QUERY/1-2623
sp|UPI0001D554F6#1/1-2600
sp|F6PVH4#1/1-2592
sp|F6YCA0#1/1-2622
sp|F7AW34#1/1-2601
sp|F6VQH4#1/1-2611
sp|E2R3I3#1/1-2629
sp|F1PFP3#1/1-2580
sp|F1SJM8#1/1-2601
sp|UPI000223546E#1/1-2619
sp|G3SKZ8#1/1-2603
sp|G1LQI6#1/1-2508
sp|UPI0000F3412A#1/1-2612
sp|G1PIE5#1/1-2620
sp|UPI0001CE2CF7#1/1-2610
sp|G1SPW0#1/1-2604
sp|UPI0001CE2CF8#1/1-2593
sp|UPI00022B71CB#1/1-2612
sp|UPI000228BF17#1/1-2588
sp|E1BAU3#1/1-2374
sp|Q3V1M1#1/1-2594
sp|G3V7S1#1/1-2597
sp|Q6WRH9#1/1-2597
sp|D4A6F8#1/1-2591
sp|F7GEA8#1/1-2611
sp|G3WYX0#1/1-2597
sp|F6UBY5#1/1-2627
sp|G1KH94#1/1-1047
sp|UPI0002038D23#1/1-1047
sp|F1NB34#1/1-1001
sp|UPI0000E80BF7#1/1-986
sp|F7CT64#1/1-978
sp|UPI0001DE89C7#1/1-977
sp|F6UBY5#2/1-919
sp|UPI00022B4307#1/1-1001
sp|G3P7F4#1/1-1012
sp|F6WKI8#2/1-569
sp|F6PVG4#2/1-569
sp|UPI00016E96F2#1/1-1104
sp|UPI00016E96F3#1/1-1104

D

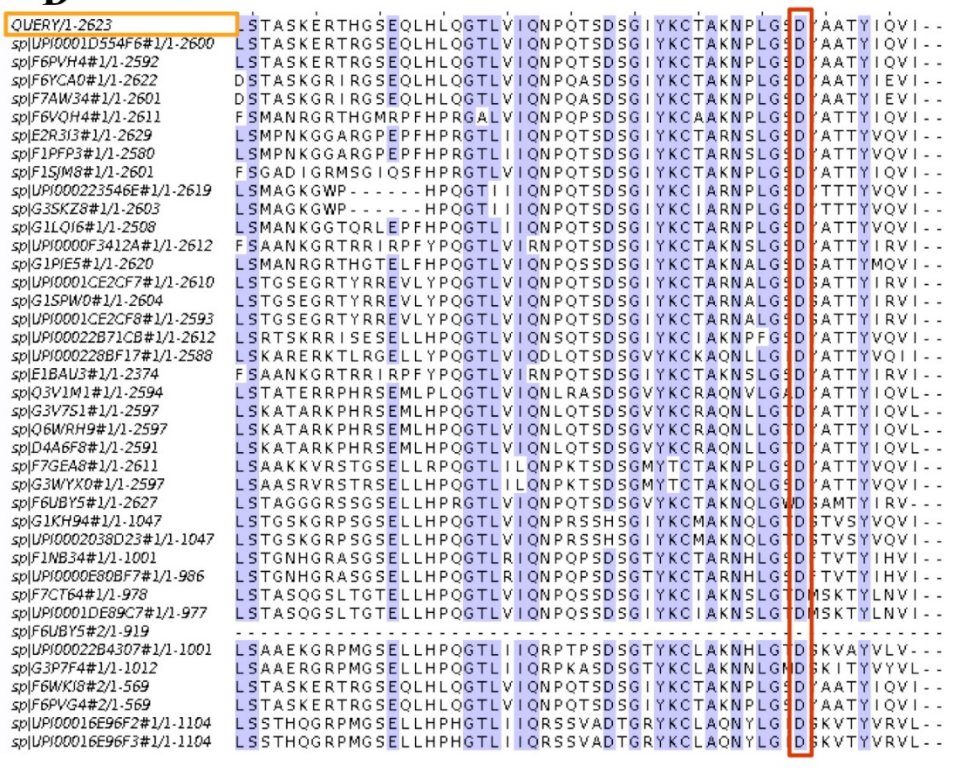

QUERY/1-2623
sp|UPI0001D554F6#1/1-2600
sp|F6PVH4#1/1-2592
sp|F6YCA0#1/1-2622
sp|F7AW34#1/1-2601
sp|F6VQH4#1/1-2611
sp|E2R3I3#1/1-2629
sp|F1PFP3#1/1-2580
sp|F1SJM8#1/1-2601
sp|UPI000223546E#1/1-2619
sp|G3SKZ8#1/1-2603
sp|G1LQI6#1/1-2508
sp|UPI0000F3412A#1/1-2612
sp|G1PIE5#1/1-2620
sp|UPI0001CE2CF7#1/1-2610
sp|G1SPW0#1/1-2604
sp|UPI0001CE2CF8#1/1-2593
sp|UPI00022B71CB#1/1-2612
sp|UPI000228BF17#1/1-2588
sp|E1BAU3#1/1-2374
sp|Q3V1M1#1/1-2594
sp|G3V7S1#1/1-2597
sp|Q6WRH9#1/1-2597
sp|D4A6F8#1/1-2591
sp|F7GEA8#1/1-2611
sp|G3WYX0#1/1-2597
sp|F6UBY5#1/1-2627
sp|G1KH94#1/1-1047
sp|UPI0002038D23#1/1-1047
sp|F1NB34#1/1-1001
sp|UPI0000E80BF7#1/1-986
sp|F7CT64#1/1-978
sp|UPI0001DE89C7#1/1-977
sp|F6UBY5#2/1-919
sp|UPI00022B4307#1/1-1001
sp|G3P7F4#1/1-1012
sp|F6WKI8#2/1-569
sp|F6PVG4#2/1-569
sp|UPI00016E96F2#1/1-1104
sp|UPI00016E96F3#1/1-1104

Figure 3.1.2 Multiple sequence alignment (m.s.a.) for the four residues harbouring mutations.

M.s.a. for residues R156 (A), E161 (B), E2264 (C) and D2614 (D). M.sa. was generated using Polyphen2 version 2.2.2, which retrieves homologues (both orthologs and paralogs, here displayed according to their Uniprot Id) using Blast+. M.s.a. is dislayed using Jalview applet. Residues are colored according to the percentage of the residues in each column that agree with the consensus sequence (percentage identity color scheme in Jalview, dark blue >80%, white < 40%). Alignments are listed in order of species, identified by abbreviation or uniprot ID.

3.2 Families with *IGSF10* variants display autosomal dominant inheritance and classical self-limited DP.

Two N-terminal variants in *IGSF10* (p.Arg156Leu and p.Glu161Lys) were identified in 20 individuals from 6 families (Fig. 3.2.1). Perfect segregation with the expected autosomal dominant pattern of inheritance was seen in all but one individual (Family3.III.3), who was found to have DP without carrying the variant. Of note, given the known association between BMI and pubertal timing, this individual was very lean (weight 13% below median weight for height) at 13 yrs (259), and his timing of puberty may have been influenced by his body mass. The two C-terminal variants (p.Glu2264Gly and p.Asp2614Asn), identified in 11 individuals from 4 families, in contrast demonstrated incomplete penetrance in family 7 and a possible *de novo* mutation in family 10 (Fig. 3.2.1).

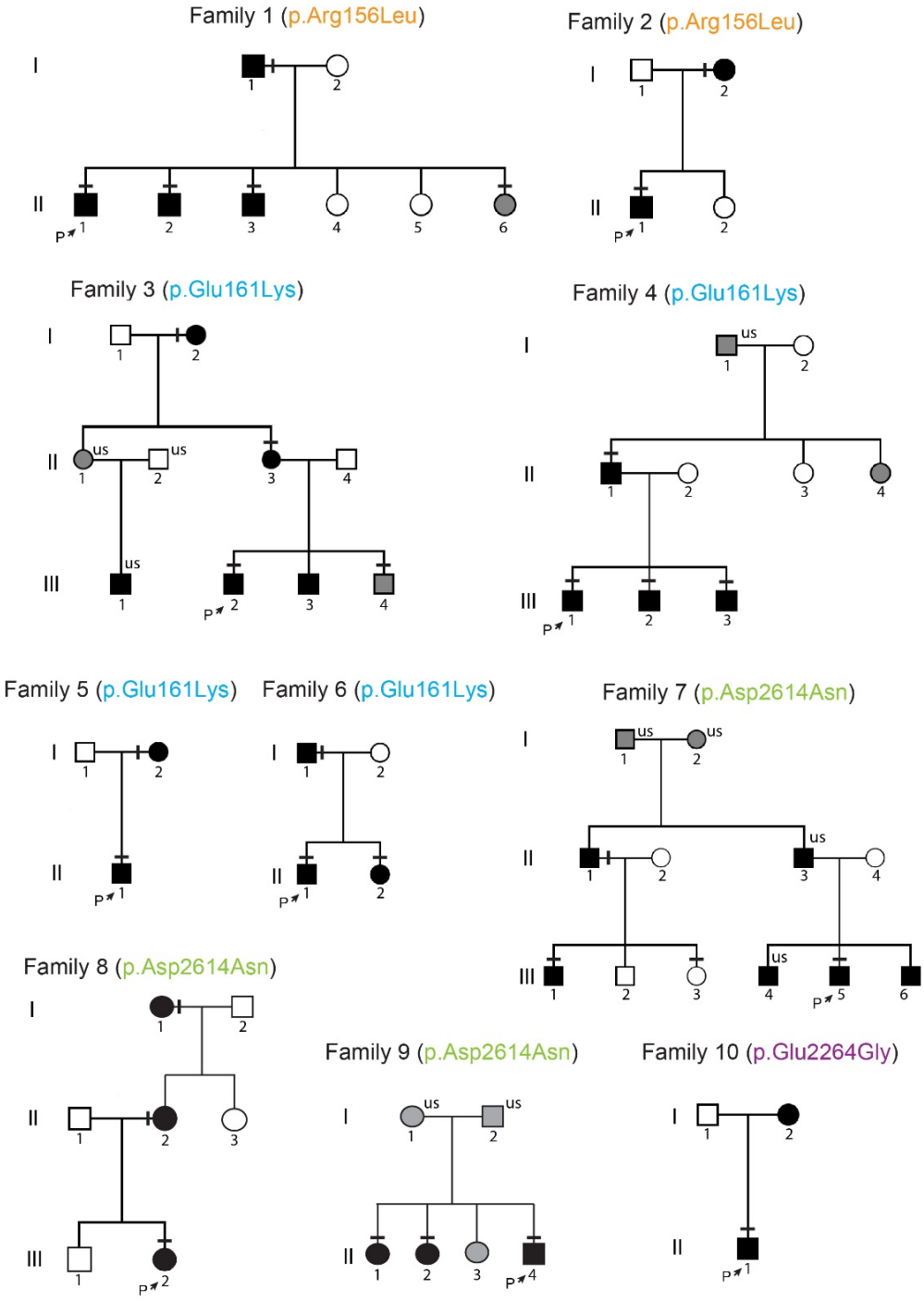


Figure 3.2.1 – Pedigrees of the families with *IGSF10* variants

Squares indicate males, circles females. Black symbols represent clinically affected, grey represent unknown phenotype, clear symbols represent unaffected individuals. The arrow with 'P' indicates the proband in each family and 'us' indicates un-sequenced due to lack of DNA. The mutation in each family is given next to the family number; a horizontal black line above an individual's symbol indicates they are heterozygous for that mutation as confirmed by either whole exome sequencing or Fluidigm array, and verified by Sanger sequencing.

Case	1.II.1	2.II.1	3.III.2	4.III.1	5.II.1	6.II.1	7.III.5	8.III.2	9.II.4	10.II.1
Clinical data at 1st assessment:										
Sex	M	M	M	M	M	M	M	F	M	M
Age (yrs)	14.76	15.5	15.5	16.01	15.69	13.66	16.14	12,18	13.55	14.94
Bone age*	12.5	12.5	13.5	13.5	13,0	12.5	13	10,0	-	-
Tvol	2,0/2,0	2,0/2,0	4,0/4,0	2,0/2,0	2,0/2,0	2,0/2,0	2,0/2,0		2.0/2.0	2.0/2.0
Ph, G or B stage	1,1	1,1	1,2	2,1	1,1	1,1	1,1	1,1	1,1	1.1
BMI	19.1	18	19.4	24.9	17,9	22.8	18.6	14.5	14.5	15
LH (IU/L) **(0.1-0.6)*** **	0.3	0.1	0.4	0.1	0.2	0.1	0.1	0.1	0.15	0.15
FSH (IU/L) **(0.1-0.9)*** **	0.45	0.1	0.5	0.3	0.6	0.35	0.6	0.2	0.5	0.3
Testost-erone (nmol/L) **(0.1-1.0)*** **	0.3	0.22	0.4	0.55	0.3	0.22	0.2	-	0.4	0.4
Estradiol (pg/mL) **(<8)*** **	-	-	-	-		-	-	<5		
Inhibin B (pg/mL) **(55-255)*** **	-	144	121	98	112	-	-		168	155
Age at:										
G2 or B2	15.2	15.6	15.5	16.5	16,10	15.21	16.5	13.94	15.21	15.4
Take-off	14.76	15.81	15.6	16.11	15,59	15.6	16.5	13.94	15.5	15.91

PHV	15.6	16.2	17	17.3	16,2	16.18	17.3	14.68	16.1	16.4
Induction of puberty:										
	yes	no	no	yes	yes	yes	yes	no	no	no
Age at start	14.76			16.11	15,59	15.21	16.5			
Duration (months)	3			6	9	9	6			
Olfaction:										
	Self-reported normal	Self-reported normal	Self-reported normal	Self-reported normal	Self-reported normal	Self-reported normal	Self-reported normal	Self-reported normal	Self-reported normal	Self-reported normal

Table 3.2.1 - Clinical and Laboratory Data of DP probands from each of the 10 families with potentially pathogenic mutations in *IGSF10*

Tvol – testicular volume in mls; Ph – pubic hair; G – genital; B – breast stage; IU – international units; G2 – genital stage 2; B2 - breast stage 2; PHV – peak height velocity. * Bone age estimated by the Greulich and Pyle method. ** baseline values. *** normal ranges for prepubertal boys given in parenthesis. Induction of puberty where indicated was with intramuscular testosterone esters.

The affected individuals from these 10 families have classical clinical and biochemical features of 'simple' DP, with delayed onset of tanner stage 2 and delayed peak height velocity (Table 3.2.1). All probands had low gonadotropins with low or undetectable sex steroids and delayed bone age at presentation.

In addition, these 10 probands displayed a typical growth pattern of self-limited DP, without compromised linear growth before puberty (Fig. 3.2.2).

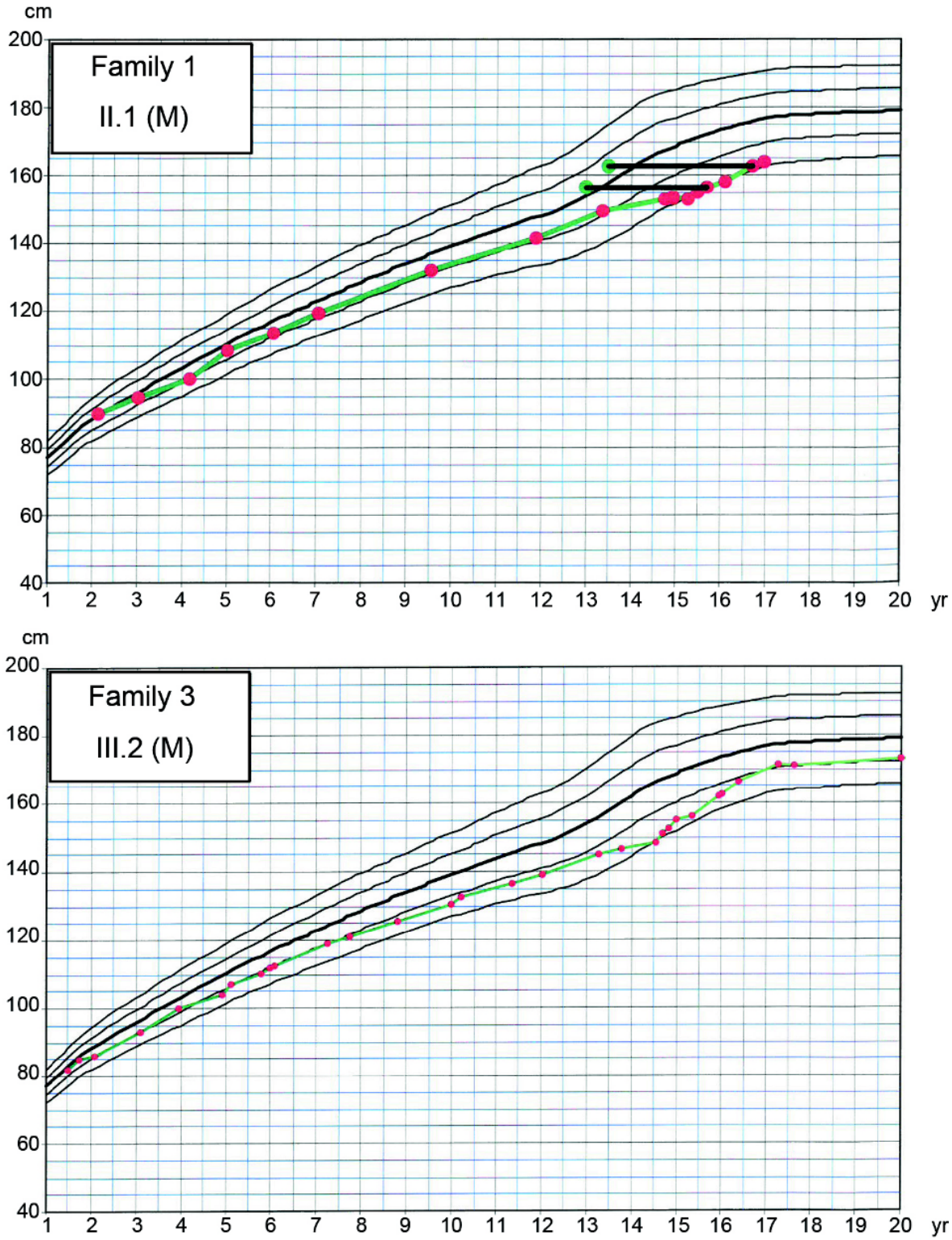


Figure 3.2.2 – Growth Charts from typical probands with *IGSF10* variants

Growth charts of 2 probands each showing typical growth patterns of self-limited DP, without compromised linear growth before puberty. Solid horizontal black lines connect green dots representing bone age to red dots at the equivalent chronological age.

Although mean height SDS was below target height, the majority of patients fell within normal limits (Table 3.2.2). Birth length, birth weight, timing of pubertal onset and adult height of those with *IGSF10* mutations were similar to those of other affected DP individuals without *IGSF10* mutations from our cohort (Table 3.2.3).

Case	Sex	Amino acid alteration	Height SDS at age 4 yrs	Height SDS at age 8/9 yrs	Target height	delta HSDS	Distance to target height at 4 yrs	Distance to target height at 8/9 yrs
1.II.1	M	p.Arg156Leu	-1.1	-1.3	-0.4	-0.2	0.7	0.9
2.II.1	M	p.Arg156Leu	-0.2	-0.4	-0.7	-0.2	-0.5	-0.3
3.III.2	M	p.Glu161Lys	-0.5	-0.8	0.7	-0.3	1.2	1.5
4.III.1	M	p.Glu161Lys	-0.4	-0.7	0.4	-0.3	0.8	1.1
5.II.1	M	p.Glu161Lys	-0.3	-0.2	0.9	0.1	1.2	1.1
6.II.1	M	p.Glu161Lys	-0.2	0	1.7	0.2	1.9	1.7
7.III.5	M	p.Asp2614Asn	-1	-1.5	-0.4	-0.5	0.6	1.1
8.III.2	F	p.Asp2614Asn	-1.9	-2.2	-1	-0.3	0.9	1.2
9.II.4	M	p.Asp2614Asn	-0.7	-1.6	-0.2	-0.9	0.5	1.4
10.II.1	M	p.Glu2264Gly	-1.8	-2	-0.6	-0.2	1.2	1.4

Table 3.2.2 – Clinical data of probands with *IGSF10* variants

Height is expressed in s.d. score (SDS) for national reference data for Finland at 4 years of age and at either 8 years for girls or 9 years for boys. Normal limits: delta HSDS <1.21, distance to target height at 4 yrs <1.76, distance to target height at 8/9 yrs <1.72(239)

Males			
	Self-limited DP patients with one of 4 *IGSF10* mutations **(mean±SD)**	**Self-limited DP patients without one of 4 *IGSF10* mutations** **(mean±SD)**	**P value**
	n=18	n=64	
Birth weight (kg)	3.6±0.6	3.7±0.4	ns
Birth length (cm)	50.6±1.9	50.9±1.2	ns
Age at G2 stage (yrs)	15.2±0.8	15.2±0.9	ns
Age at "take off" (yrs)	15.0±1.3	14.9±0.7	ns
Age at PHV (yrs)	16.3±1.2	16.1±1.0	ns
Adult height (cm)	177.0±5.9	179.3±7.0	ns
Females			
	Self-limited DP patients with one of 4 *IGSF10* mutations **(mean±SD)**	**Self-limited DP patients without one of 4 *IGSF10* mutations** **(mean±SD)**	**P value**
	n=10	n=125	
Birth weight (kg)	3.7±0.2	3.6±0.5	ns
Birth length (cm)	50.3±2.3	50.0±2.2	ns
Age at B2 stage (yrs)	13.5±1.0	13.8±1.0	ns
Age at "take off" (yrs)	13.3±0.7	12.5±1.1	ns
Age at PHV (yrs)	14.5±0.7	13.4±1.0	ns
Adult height (cm)	168.5±5.4	165.1±6.5	ns

Table 3.2.3 – Auxological Features and Pubertal Timing of self-limited DP patients from the cohort with identified variants in *IGSF10.*

Tables are divided into male and female and compare auxological data and pubertal timing of those self-limited DP patients with mutations in *IGSF10* with those without mutations in *IGSF10.* All comparisons were found to be non-significant (ns) by Student's *t* test (2-tailed); G2 – Tanner genital stage 2; B2 – Tanner breast stage 2; "take off" – age at start of pubertal growth spurt; PHV – peak height velocity

3.3 *In silico* Analysis

IGSF10 (immunoglobulin superfamily member 10) is a gene with thus far unclear function, mutations in which have not previously been associated with disease in humans. The IGSF10 protein has not previously been modelled by crystallography, so we therefore performed *in silico* analysis (Fig. 3.3.1). *In silico* analysis of the protein reveals five defined regions. Region I, in which the two N-terminal variants identified are located, contains leucine-rich repeats (LRR) 1-7 flanked by a LRR N-terminal (LRR Nt) and C-terminal (LRR Ct) cap. Region II is structurally disordered. Region III contains two immunoglobulin-like beta sandwich (Ig-like) domains. Region IV is structurally disordered. Region V, in which the two C-terminal variants are located, contains 10 further Ig-like domains (Fig. 3.3.1). No clear evidence for a predicted transmembrane domain was provided by the *in silico* analysis. However, published data gives evidence for a putative cleavage site with secretion of the N-terminal portion (246).

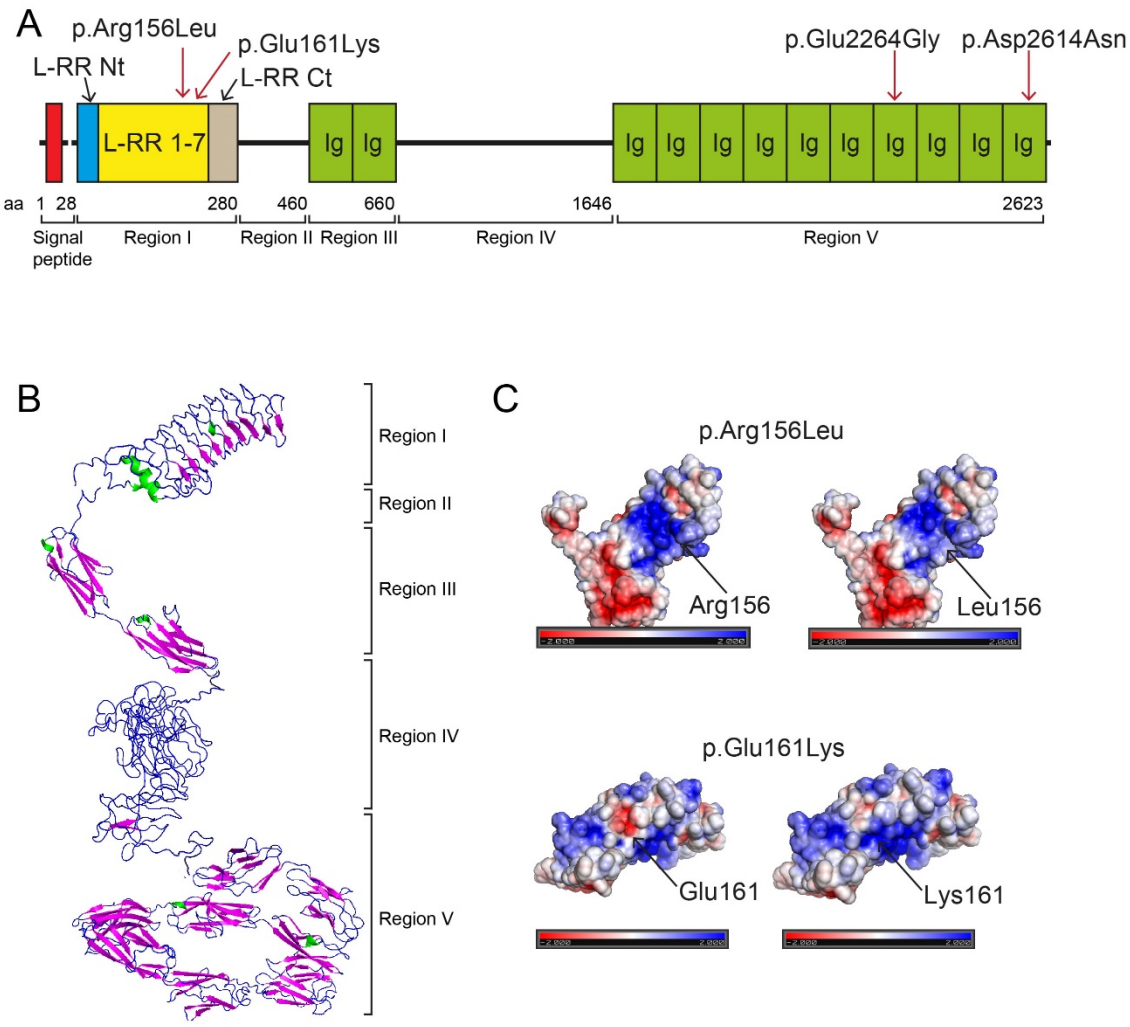


Figure 3.3.1. IGSF10 protein structure and position of identified mutations.

Panel A: IGSF10 domains and variants identified in the study. Region I contains leucine-rich repeats (LRR) 1-7 flanked by a LRR N-terminal (LRR Nt) and C-terminal (LRR Ct) cap. Region II is structurally disordered. Region III contains two Ig-like domains (Ig). Region IV is structurally disordered. Region V contains 10 Ig-like domains (Ig).

Panel B: Protein tertiary structure as predicted by *in silico* analysis. Panel C: Electrostatic maps showing p.Arg156Leu and p.Glu161Lys variants: p.Arg156Leu variant - the electrostatic map shows a large positively charged region (displayed in blue) on the protein concave surface. Substitution of Arg with the neutrally charged Leu, results in loss of part of the dark blue area; p.Glu161Lys - the electrostatic map shows a negatively charged patch (displayed in red) on the protein convex surface. Substitution of Glu with Lys, results in loss of charge represented by loss of the dark red area.

Two variants identified in our patients (p.Arg156Leu, p.Glu161Lys) are located in region I within LRR domains, and two (p.Glu2264Gly, p.Asp2614Asn) within Ig-like domains in region V at the C-terminal of the protein (Fig. 3.3.1A). Structural analysis reveals that Arg156 is part of the large positively charged region on the concave face of the LRR domain, which may be a protein-protein interaction site (Fig. 3.3.2 and Fig. 3.3.3).

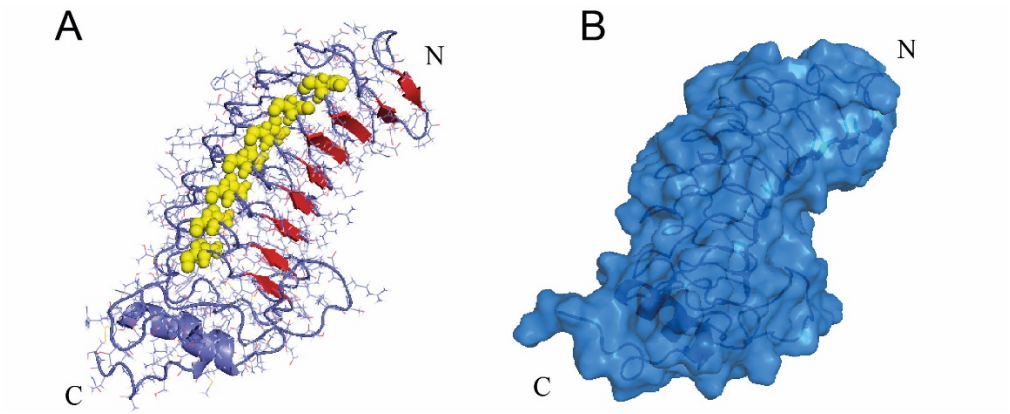


Figure 3.3.2 - Tertiary structure of L-RR Region I. Panel A: Beta strands are displayed in red and the asparagine ladder in yellow. Panel B: the 'arc' shape of L-RR region I.

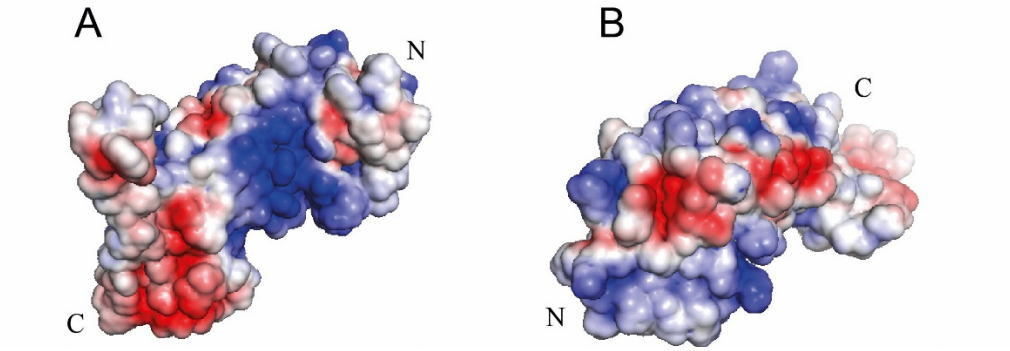


Figure 3.3.3 - Electrostatic map of L-RR Region I. Visualisation of this region's electrostatic potential demonstrates the presence of a large positively charged patch on the inner concave surface (panel A) and several interconnected negatively charged patches on the outer surface (panel B). The latter is likely to represent a protein-protein interaction site.

Substitution of the positively charged Arg with the neutral Leu may disrupt this interaction site. (Fig. 3.3.1C). Glu161 is part of a negatively charged patch on the convex face, and substitution with a positively charged Lys could also result in disruption of this putative protein interaction site (Fig. 3.3.1C). The C-terminal mutations p.Glu2264Gly and p.Asp2614Asn are predicted to disrupt a site of domain-domain interaction and weaken the formation of an intradomain salt bridge respectively (Fig. 3.3.4 and Fig. 3.3.5).

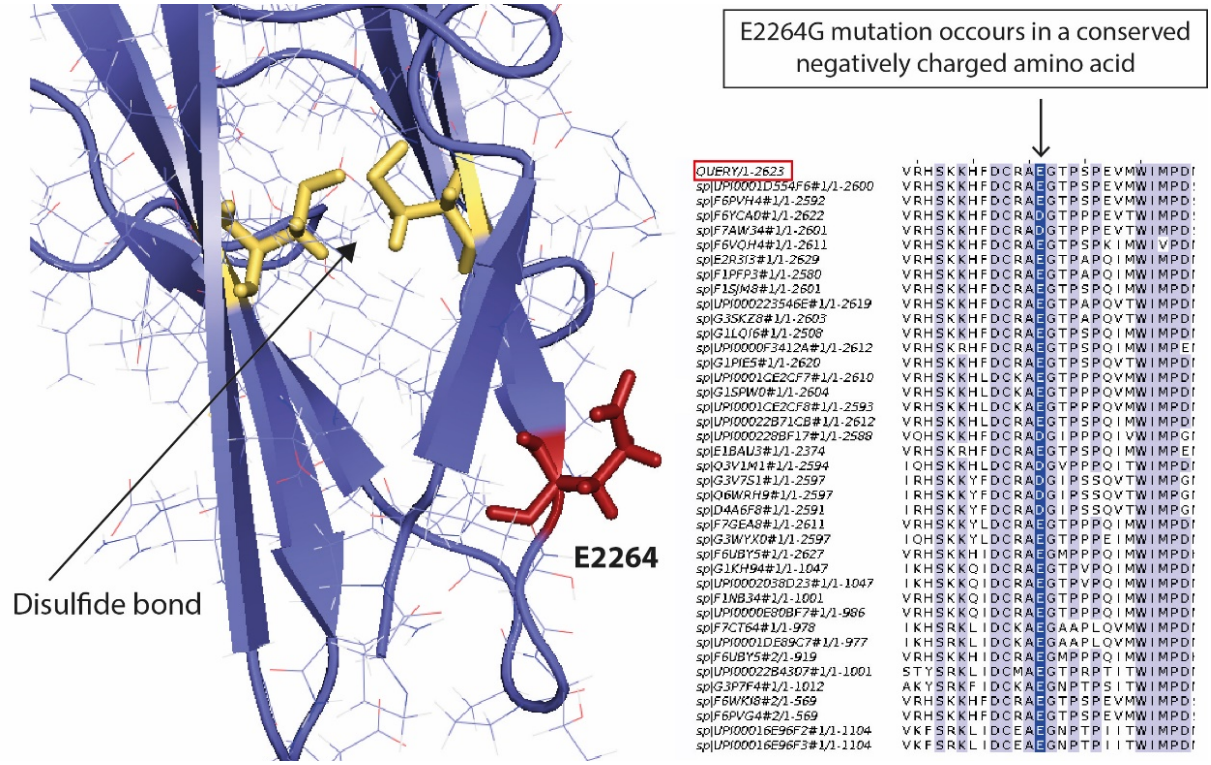


Figure 3.3.4 - Mutation p.Glu2264Gly (E2264G).

Amino acid Glu2264 (E2264, in red) is localized on the surface of the Ig-like domain. Multiple sequence alignment shows that this position is occupied by negatively charged amino acids (the Uniprot Id of each homologue is indicated). Charge conservation amongst homologues and structural localization suggest that this amino acid may be involved in domain-domain interaction and substitution with the neutral Gly may well disrupt this possible interaction site. The disulphide bond in this Ig-like domain is shown in yellow.

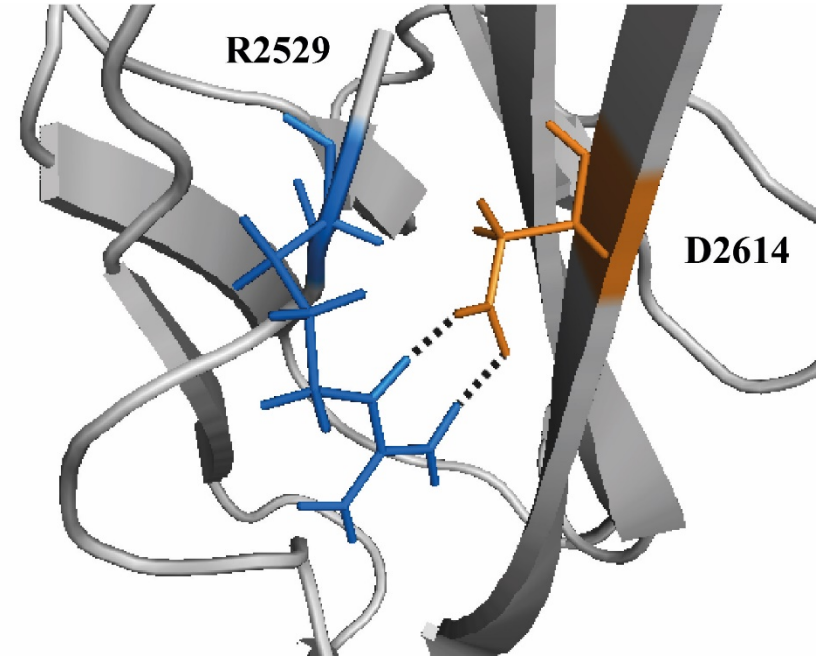


Figure 3.3.5 - Mutation p.Asp2614Asn (D2616N).

Asp2614 (D2614, in orange) is predicted to form an intradomain a salt bridge with Arg2529 (R2529, in blue). Substitution of Asp2614 (D) with Asn (N) may weaken this interaction and destabilize the domain structure.

3.4 IGSF10 N-terminal mutant proteins display pathogenic features with failure of extracellular secretion.

The two N-terminal variants identified (p.Arg156Leu, p.Glu161Lys) are located in region I within LRR domains (Fig. 3.3.1A). Both WT and mutant N-terminal protein fragments (668 aa in length) were expressed in HEK293 cells as demonstrated by western blotting. However, whilst the GFP-tagged WT protein was detected in the conditioned media, neither mutant protein could be detected in their respective conditioned media. Moreover, a significant increase in mutant protein was detected in cell lysates, suggesting intracellular retention of these two mutants (Fig. 3.4.1).

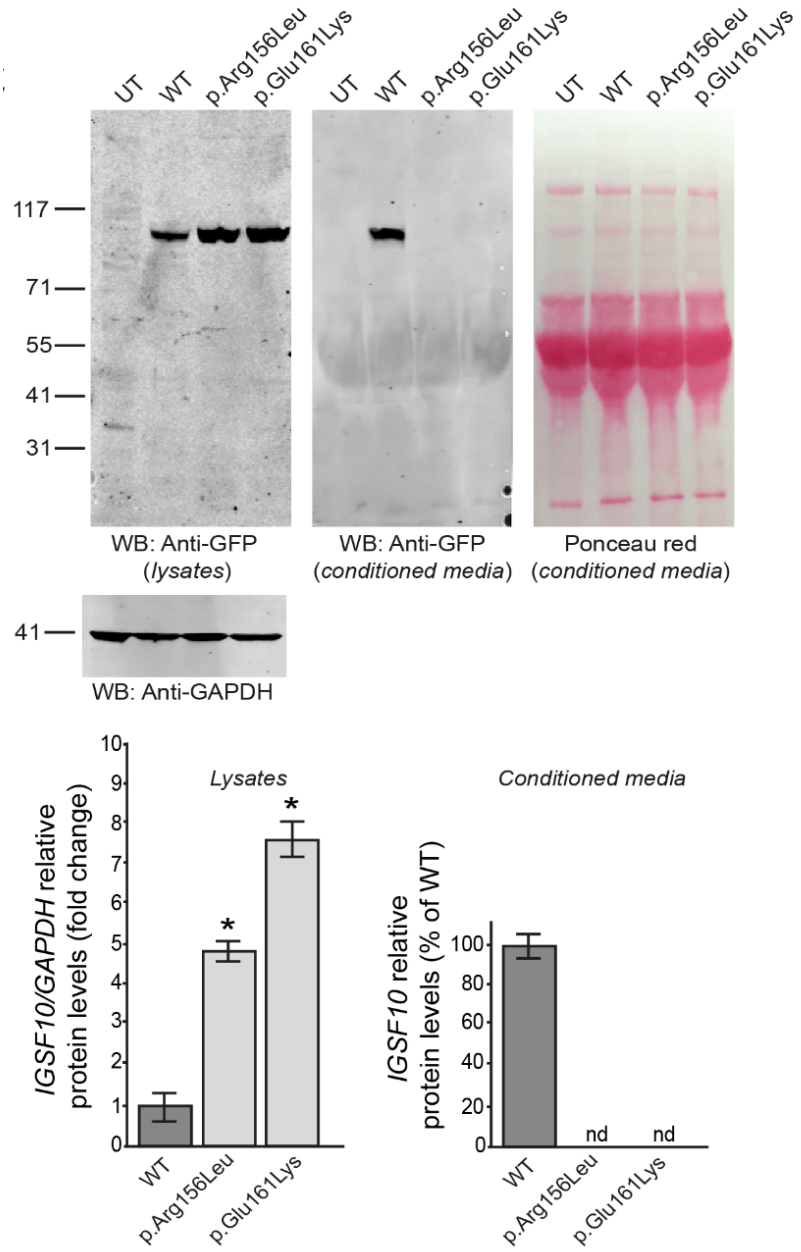


Figure 3.4.1 - Biological consequences of the 2 identified N-terminal mutations.

Both WT and mutant N-terminal protein fragments (p.Arg156Leu and p.Glu161Lys) were expressed in HEK293 cells as demonstrated by western blotting. The GFP-tagged protein products of both were not detected in the conditioned media of mammalian cells, as compared to wild type (WT), and appear to be retained in the intracellular compartment (mean ± s.e.m.; *n* = 3). Ponceau red staining is shown to demonstrate equal protein loading for conditioned media. UT – untransfected negative control; nd – not detected; *: $p<0.05$ by 2-tailed t-test. It has to date not been possible to test cytoplasmic retention for the two C-terminal mutations due to difficulty expressing the full length or C-terminal protein fragment in mammalian cells.

3.5 *IGSF10* tissue expression by RT-qPCR

Tissue expression of *IGSF10* was first analysed in a human adult tissue panel via RT-qPCR. Highest expression was seen in ovary, colon and lung (Fig. 3.5.1).

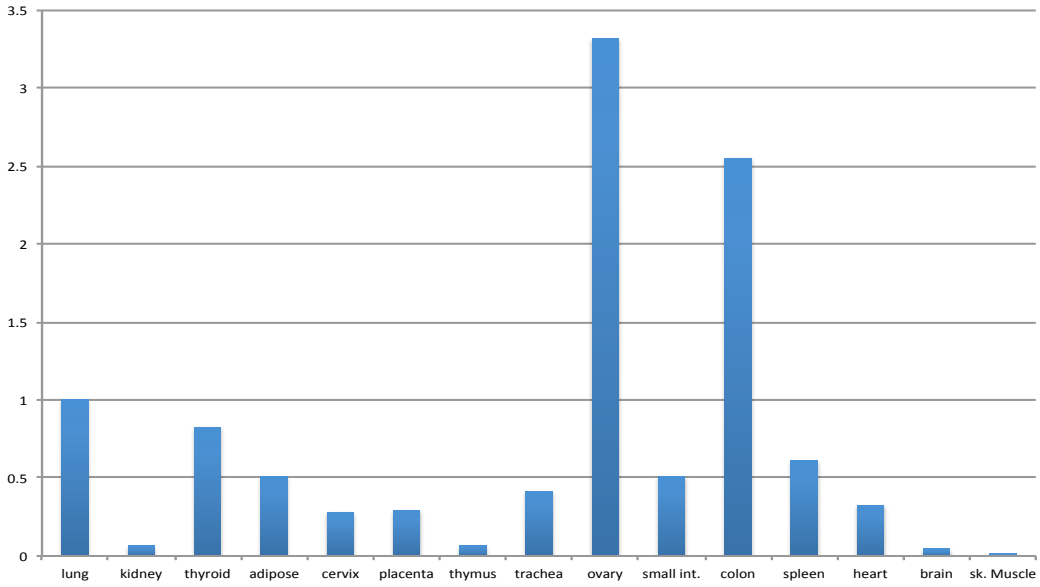


Figure 3.5.1 – Expression of *IGSF10* in a human adult tissue panel by RT-qPCR. Expression is shown relative to *GAPDH*.

However, published data demonstrated the expression of *Igsf10* mRNA in the developing mouse embryo in the mesenchymal cells around the nasal placode and in the branchial arches (260). In view of the importance of the migrational journey of GnRH neurons to the normal development of the HPG axis, and that several chemokines and cytokines are known to regulate this migration through the nasal mesenchyme (1), I went on to investigate this embryonic expression further.

3.6 Tissue expression studies localize *Igsf10* mRNA expression to the spatial and temporal window of GnRH neuronal migration.

The expression of *Igsf10* mRNA in the nasal region of mouse embryos was analysed from embryonic day (E) 10.5 to E17.5. During this developmental window GnRH neurons emerge from the nasal placode (around E10.5) and then migrate towards the basal forebrain and hypothalamus (E12.5 to E17.5)(1). *Igsf10* mRNA expression was below detectability at E10.5 (Fig. 3.6.1, panel A). At E12.5 *Igsf10* was prominently expressed in the nasal mesenchyme (NM) with a decreasing gradient of expression from the area surrounding the vomeronasal organ (VNO) towards the nasal forebrain junction (NFJ), and absent in the VNO and olfactory epithelium (OE) (Fig. 3.6.1, panel B). At E12.5, GnRH neurons are exiting the VNO and migrating into an *Igsf10* strongly positive cell milieu in the NM (Fig. 3.6.1, panels B and C). At E14.5 GnRH cells and peripherin-positive olfactory axons are navigating amongst the *Igsf10* positive NM cells (Fig. 3.6.1, panels D, G and H).

Cxcl12, one of a plethora of molecules known to provide directional cues to GnRH neurons (247), is also expressed in the NM, although with an opposite gradient compared to *Igsf10* (Fig. 3.6.1, panel I). At E17.5 GnRH neurons are mainly located in the medial preoptic area (MPOA, Fig. 3.6.1, panel J). *Igsf10* signal was not detected in the hypothalamus at E17.5 or in peri-pubertal adult mice (data not shown). A similar expression pattern of *IGSF10* was detected in the nasal area of human embryos at 11 post-conceptual weeks (pcw) (Fig. 3.6.1, panels K-O).

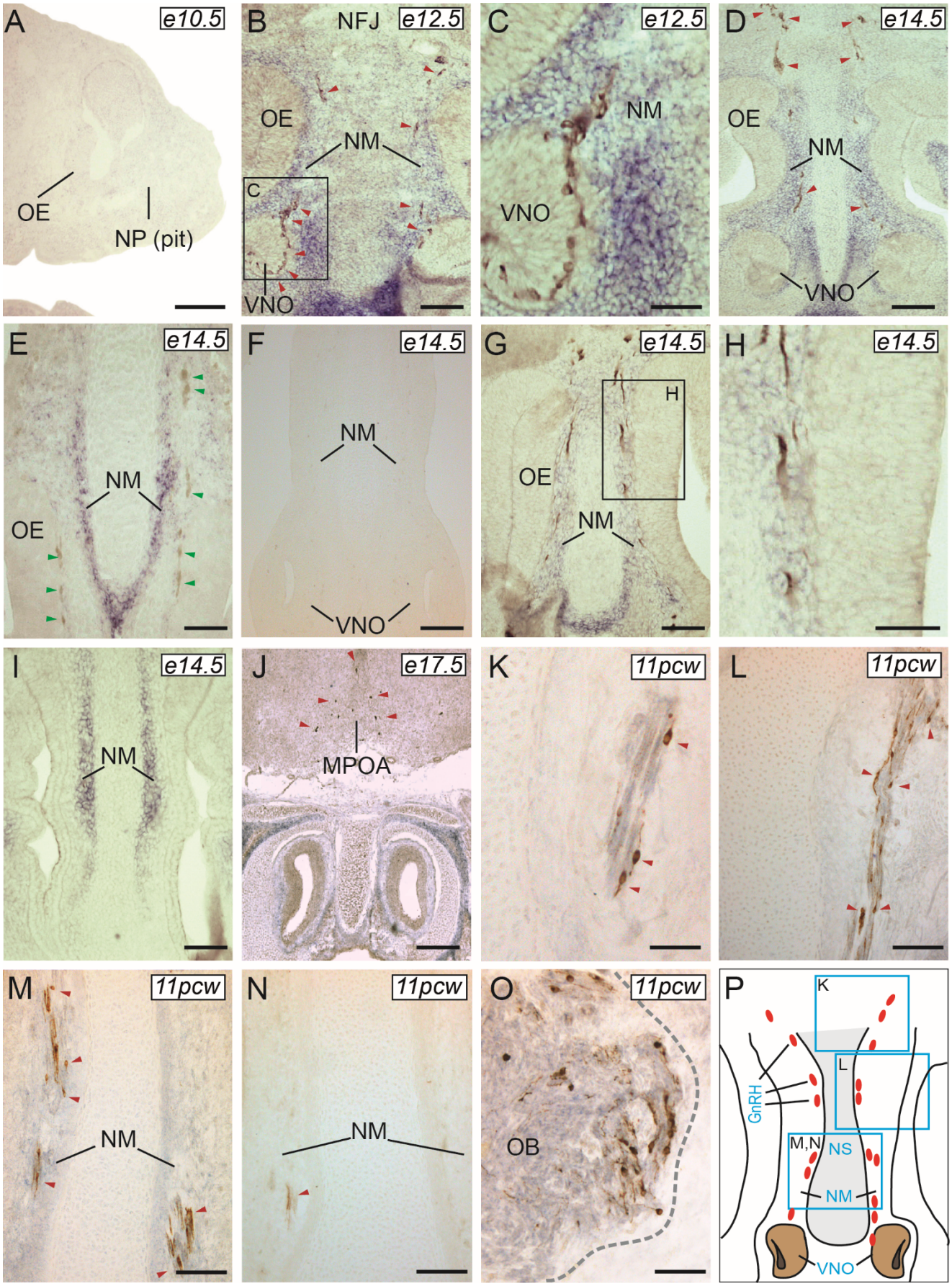

A
e10.5
OE
NP (pit)
B
NFJ
e12.5
OE
NM
C
VNO
C
e12.5
NM
VNO
D
e14.5
OE
NM
VNO
E
e14.5
NM
OE
F
e14.5
NM
VNO
G
e14.5
H
OE
NM
H
e14.5
I
e14.5
NM
J
e17.5
MPOA
K
11pcw
L
11pcw
M
11pcw
NM
N
11pcw
NM
O
11pcw
OB
P
K
L
GnRH
M,N
NS
NM
VNO

Figure 3.6.1 - Expression pattern of *Igsf10* mRNA in mouse and human developing brain

Igsf10 expression was not observed at E10.5 (A, sagittal section) but readily detectable at E12.5, in the NM (B and C, frontal sections) and E14.5, along the migratory path of GnRH neurons (D, frontal section). Igsf10 positive cells were negative for Isolectin b4 (marker of vasculature) at all developmental stages (E, at E14.5). Incubation with the sense probe resulted in no signal at all stages (E14.5 shown in F). Peripherin-positive cells olfactory axons were extending within the Igsf10-positive NM (G and H). Cxcl12 mRNA expression is shown in I (E14.5, frontal section). GnRH neurons are located in the MPOA by E17.5 (J, frontal section) in an Igsf10 negative area. In human 11pcw brains, IGSF10 expression pattern was similar to that observed in mouse, with GnRH neurons interspersed in an IGSF10-positive NM (K-N, frontal sections, see also schematic in P). IGSF10 positive cells were also detected in the olfactory bulb (OB, panel O). Sense probe resulted in no specific signal (panel N). GnRH neurons are shown by red arrowheads and Isolectin b4 by green arrowheads. NS: nasal septum.

3.7 *In vitro* functional analyses of *Igsf10*

To gain insight into the functional role of *Igsf10* in the migration of GnRH neurons we took advantage of a model of immortalized but migrating mouse GnRH cells, named GN11 cells (261). These cells express neuronal markers and retain many features of immature GnRH-secreting neurons (262, 263), including a strong chemo-migratory response *in vitro* (264-266). Co-culture experiments of GN11 aggregates placed on confluent NIH3T3 monolayers were performed. NIH3T3 cells, derived from a mouse embryonic fibroblast cell-line, express high levels of endogenous *Igsf10.* The NIH3T3 cells were treated with scrambled- or Igsf10-shRNAs, the latter leading to highly reduced levels of *Igsf10* expression (Fig. 3.7.1, A). Migration of GN11 into Igsf10-shRNA NIH3T3 cells was significantly reduced compared to that into scrambled-shRNA NIH3T3 (2-tailed t-test, n=3 for both Igsf10-shRNA and scr-shRNA with a minimum of 8 micromass replicates in each group, p=0.02422) (Fig. 3.7.1, B and C, analysis in D). Statistical significance was evaluated by ANOVA and Student's t test. $P < 0.05$ (*) and $p < 0.01$ (**) were taken to be significant. Data are presented as mean ± standard

error of mean and expressed as percentage reduction of migration, taking untreated or scrambled controls as 100%.

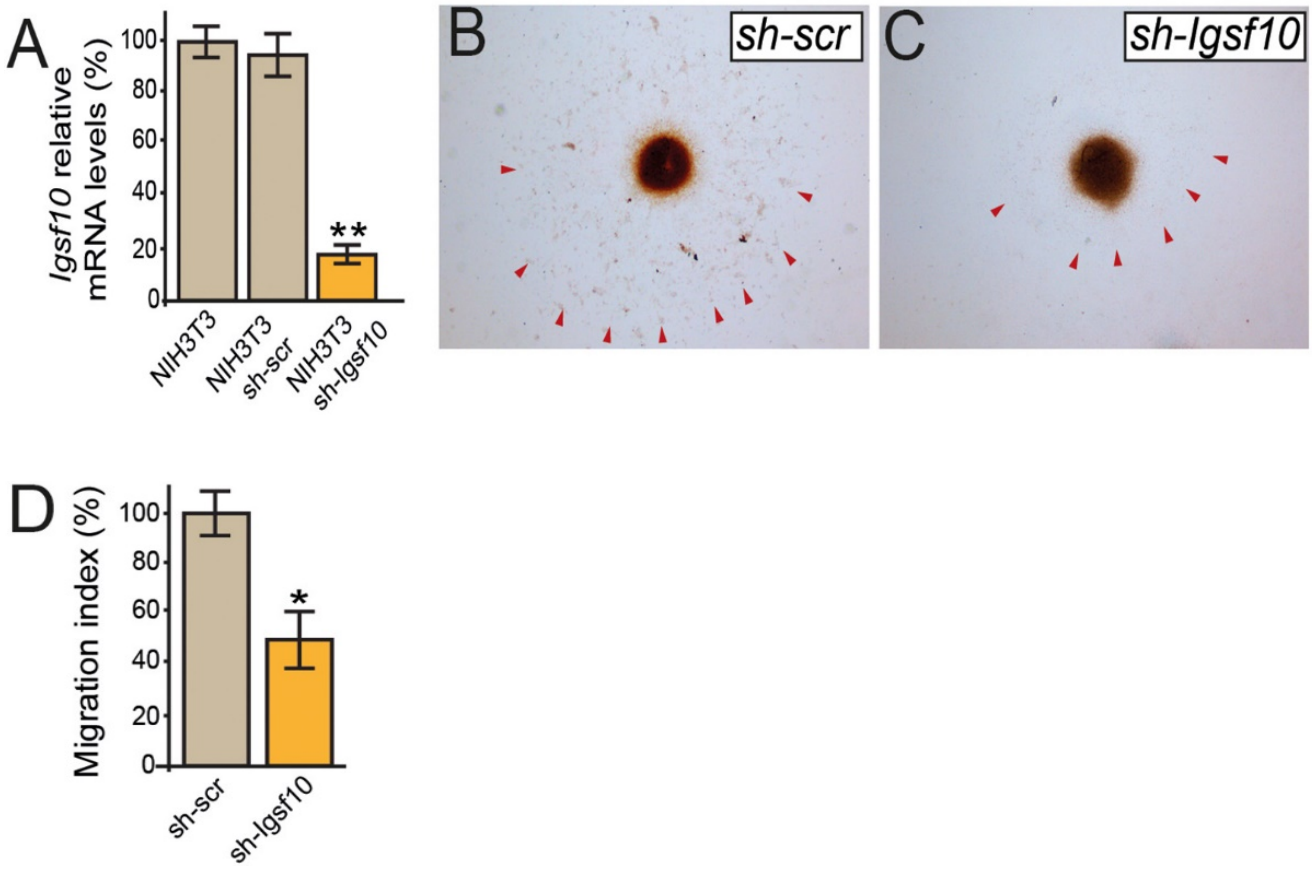


Figure 3.7.1 – Effect of *Igsf10* knockdown on GnRH neuronal migration

Panel A: Levels of *Igsf10* expression in native NIH3T3 cells, and cells stably expressing a scrambled (sh-scr) or *Igsf10* (sh-Igsf10) shRNA.

Panels B and C: migration of GN11 cells from aggregates into NIH3T3 sh-scr (B) or sh-Igsf10 (C).

Panel D: Analysis of the migration index. *: $p<0.05$, **: $p<0.01$.

3.8 *Igsf10* knock-down *in vivo* results in perturbed migration and failed neurite extension of GnRH3 neurons in zebrafish embryos

A transgenic zebrafish line, *Tg(gnrh3:EGFP)*, was used to visualize GnRH3 neurons and their projections (249). *Igsf10* expression was clearly detectable by *in situ* hybridisation in zebrafish embryos at 48 hpf in the head, adjacent to and surrounding the staining for *Gnrh3* expression, which at 48 hpf is seen as two discrete points (Fig. 3.8.1) However, *Igsf10* expression was at low resolution and so it was difficult to make a functional interpretation of the role Igsf10 might be playing in zebrafish embryos via ISH studies alone.

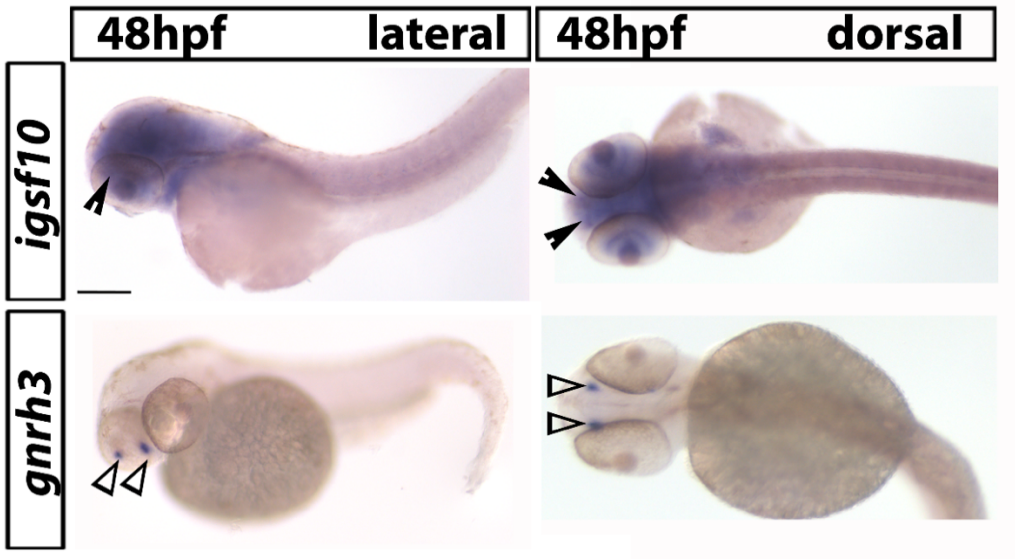


Figure 3.8.1 –*Igsf10* expression in zebrafish embryos at 48 hours post fertilisation (hpf). Igsf10 expression is shown by dark blue/purple staining in the top panel. Gnrh3 expression is shown by dark blue/purple staining in the bottom panel.

RT-PCR analysis indicated high and relatively equal expression of *Igsf10* mRNA at all time-points tested during embryogenesis, from 48 hours post-fertilisation (hpf) onward (Fig. 3.8.2, A). Morpholino-modified antisense oligonucleotides (MOs) were used to repress

Igsf10 mRNA to assess the effect of *Igsf10* knockdown on the development of the GnRH3 system. Time-lapse analysis of MO-injected *Tg(gnrh3:EGFP)* embryos is shown in Fig. 3.8.2, C. At 48 hpf, GnRH3 neurons are normally seen as bilateral dots in the olfactory organ–olfactory bulb boundary. Over the following days their projections extend caudally through the telencephalon to the hypothalamus. The strongest effect of morpholino injections was observed at 48hrs. At this time-point, the percentage of embryos showing an abnormal GnRH3-neuron phenotype was higher in Igsf10 Splice-site-MO (Sp-MO) injected embryos compared to relative controls, injected either with a mispair morpholino (Cntrl-MO) or uninjected (mean ± standard error of mean: Sp-Mo 33% ± 3.6 vs. Cntrl-MO 11.7% ± 1.5 vs. uninjected 2.8 ± 1.3; 1-way ANOVA, n=201/160/156 for Sp-MO/Ctrl-MO/uninjected, p=0.000009, Fig. 3.8.2, B).

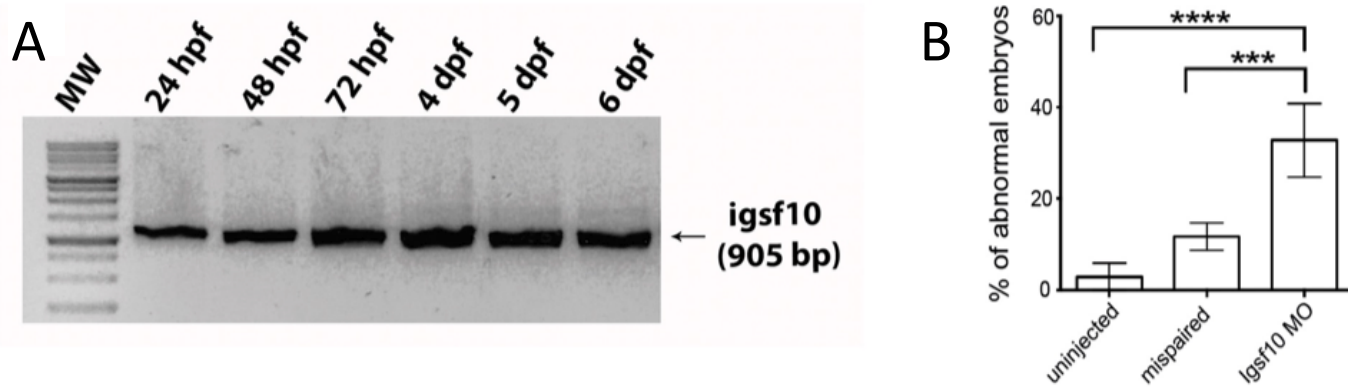
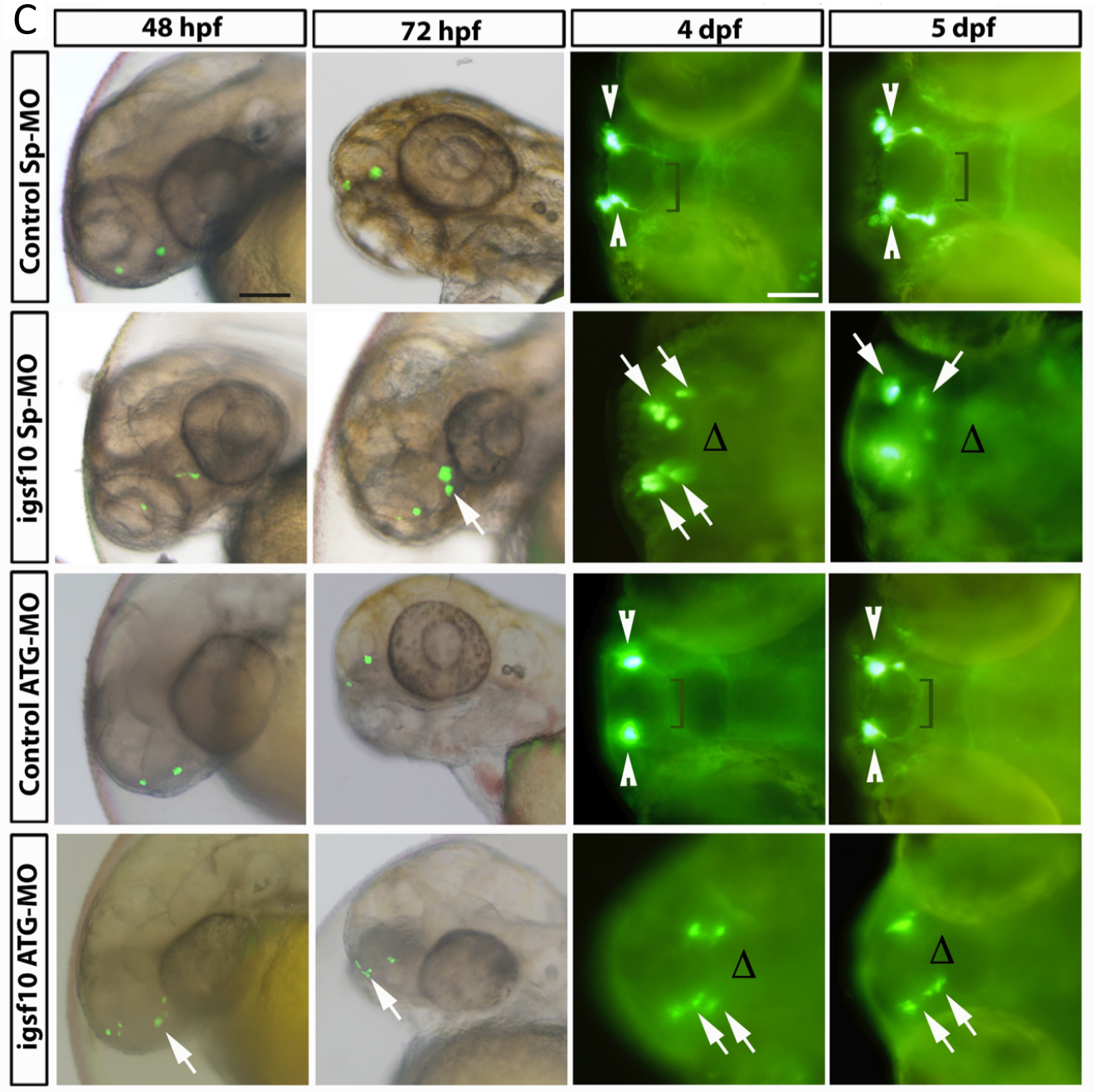


Figure 3.8.2 – Effect of *Igsf10* knockdown on GnRH neuronal migration

Panel A: RT-PCR analysis in zebrafish of total embryos and larvae showing the expression of *Igsf10* mRNA at different time points. Expected band size: 905 bp

Panel B: Effect of Igsf10 Splice-site morpholino injection observed at 48hrs. Quantification of the percentage of embryos showing an abnormal GnRH3-neuron phenotype was significantly higher in Igsf10 Splice-site-MO (Sp-MO) injected embryos compared to relative controls. ***: p<0.001, ****: p<0.0001.



Panel C: Representative time-lapse analysis of mispair control (Control Sp-MO and Control ATG-MO) and Igsf10 morpholinos (Igsf10 Sp-MO and Igsf10 ATG-MO) injected *Tg(gnrh3:EGFP)* embryos and larvae, at different time-points. Images show lateral view of live larvae head (48 and 72 hpf) and dorsal view (4 and 5 dpf) of live larvae head; anterior is left. White arrowheads indicate normal GnRH3 neuron clusters in the olfactory area of control embryos. Brackets show the extension of the projections towards the hypothalamus. White arrows indicate examples of abnormal GnRH3 neurons scattered in the olfactory area of morphants accompanied by lack of projections toward the hypothalamus, indicated with Δ (hpf – hours post fertilisation; dpf – days post fertilisation.) Scale bar, 100 um (panels at 48 and 72hpf in F), 150 um (panels at 4 and 5dpf in F). Experimental work and images by Dr. Valentina Andre.

A similar effect was observed with the embryos injected with Igsf10 ATG-MO. *Igsf10* knockdown affected both the guidance and the axonal outgrowth of GnRH3 neurons, which were unable to form compact clusters or extend projections to the hypothalamus.

3.9 Cross-Phenotype Analysis

The ReproGen Consortium analysis did not identify any variants in *IGSF10* that reached statistical significance in their AAM data, suggesting that with the current threshold for identifying statistically significant variants by GWAS, variants in *IGSF10* do not have a strong influence on the timing of puberty in the general population. However, we plan to repeat this analysis with future meta-analyses of GWAS of AAM by the ReproGen consortium, as large numbers of women continue to be recruited to these studies. Additionally, studies of age of voice breaking in men are also being included in these analyses (203).

3.10 Loss-of-function mutation in *IGSF10* in 2 patients with functional hypogonadotropic hypogonadism.

To explore the possible role of mutations in *IGSF10* in conditions of GnRH deficiency, we carried out targeted exome sequencing of *IGSF10* in an adult cohort of 334 patients with HH due to Kallmann syndrome, idiopathic HH or functional hypogonadism (hypothalamic amenorrhoea (HA) or HA-equivalent). This investigation showed that 10.2% of these patients carried a rare, predicted damaging variant in *IGSF10* (Table 3.10.1).

HH cohort (n=334)			ExAC database (n=32200)			P value
Total number of rare and predicted damaging variant alleles*	Allele Frequency	Total Prevalence	Total number of rare and predicted damaging variant alleles*	Allele Frequency	Total Prevalence	
34	5.1%	10.2%	2314	3.5%	7%	0.0329

Table 3.10.1 - Frequency of *IGSF10* variants in the hypogonadotropic hypogonadism (HH) cohort.

Minor allele frequency (MAF) data for the European control subjects was retrieved from the ExAC Browser (Exome Aggregation Consortium (ExAC), Cambridge, MA: http://exac.broadinstitute.org, accessed October 2015). P value calculated by Fisher's exact test. *rare variants with MAF <2.5% and predicted damaging by both SIFT and Polyphen2

Three loss-of-function variants (NM_178822: c.C352T: (rs142596318) p.R118*, NM_178822: c.G4804T: (rs79363433) p.E1602* and NM_178822: c.7353_7354insATCA: (rs570110855) p.L2452fs) were identified in a total of 5 patients (Table 3.10.2), and 13 missense variants were identified in 29 patients from our HH cohort, all in the heterozygous state. In particular, all three of the loss-of-function variants were enriched in our HH cohort as compared to the ExAC database.

Two patients (out of 14 patients from this HH cohort with functional hypogonadism) were identified as heterozygous for a shared loss-of-function mutation in *IGSF10* (NM_178822: c.7353_7354insATCA: (rs570110855) p.L2452fs) (Table 3.10.3). This variant is predicted to be deleterious with a high degree of confidence, with the expected loss of the last two Ig-like domains of the IGSF10 protein. It is a rare variant in the general population, with a

minor allele frequency in the ExAC database of 0.01%. Both patients carrying the variant had adult-onset functional hypogonadism associated with environmental stressors. Patient 1 had a history of secondary amenorrhoea induced by excessive exercise, and presented due to failure to achieve spontaneous pregnancy aged 31yrs. Patient 2 presented with adult-onset hypogonadism secondary to excessive weight loss and a sub-clinical eating disorder. Neither patient had a family history of HH, and both had normal pituitary imaging and normosmia.

Patient	Phenotype	Variant	Segregation
1	Male Kallmann syndrome	p.Arg118*	Non-complete
2	Male Partial hypogonadotropic hypogonadism, cryptorchidism and micropenis	p.Glu1602*	*De novo* variant in proband, parents unaffected
3	Female Partial hypogonadotropic hypogonadism with normal olfactory bulbs	p.Glu1602*	Relatives unsequenced
4	Female Hypothalamic amenorrhoea (HA)	p.Leu2452fs	Relatives unsequenced
5	Male-equivalent HA	p.Leu2452fs	Relatives unsequenced

Table 3.10.1 – *IGSF10* predicted loss-of-function variants identified in the HH/ HA cohort.

Characteristic	Patient 1	Patient 2
Clinical Characteristics		
Gender	Female	Male
Age (yr)		
At menarche	Unknown	-
At diagnosis	25	29
BMI at diagnosis	25 (low body fat to muscle ratio)	18
Predisposing factors		
Weight loss	No	Yes
Excessive exercise	Yes	No
Subclinical eating disorder	No	Yes
Fertility Status	Failed attempt at conception off treatment; undergoing gonadotropin therapy	No attempt at conception
Family history of hypothalamic amenorrhea	No	No
Radiological Findings		
MRI	Normal pituitary	Normal pituitary
Biochemical Findings		
LH (U/L)	5.0	0.6
FSH (U/L)	5.9	4.1
Estradiol (E_2) (pg/mL)	<15	-
Testosterone (ng/mL)	<0.3	0.8
Prolactin (mU/L)	129	111
TFTs	Within normal range	Within normal range
AMH	40.3	-

Table 3.10.3 - Clinical characteristics of two patients with hypothalamic amenorrhea (HA) and HA-equivalent with a shared rare variant in *IGSF10* (NM_178822: c.7353_7354insATCA: (rs570110855) p.L2452fs). The body-mass index (BMI) is the weight in kilograms divided by the square of the height in metres; LH- luteinising hormone; FSH – follicular stimulating hormone; TFTs – thyroid function tests, comprising of free T4 and thyroid stimulating hormone; AMH – anti-mullerian hormone

3.11 Discussion

In this study, our findings indicate a role for *IGSF10* in the migration of GnRH neurons, and highlight two pathogenic mutations in *IGSF10* as the causal factor for DP in 6 unrelated families. We have identified an additional two rare variants of unknown significance in 4 further families. These two additional variants lie within Ig-like domains and are anticipated to either impair protein-protein interaction or weaken domain structure.

IGSF10, a gene of previously unclear function, is a member of the immunoglobulin superfamily. Loss-of-function mutations in another member of this superfamily, *IGSF1*, were recently identified in patients with X-linked central hypothyroidism (253). Notably, male patients with *IGSF1* mutations have a late increase in testosterone levels with a delayed pubertal growth spurt.

Our functional work specifies a likely role for *IGSF10* in the early migration of GnRH neurons. The development of the hypothalamic-pituitary-gonadal axis is exceptional in that GnRH neurosecretory neurons develop in metazoan embryos outside of the central nervous system. Immature GnRH precursor neurons are first detectable in the olfactory placode in the nose from an early embryological stage (E10.5 in mice) and then begin a complex migration through the forebrain into the hypothalamus and preoptic areas (1, 55).

An intact GnRH neurosecretory network is necessary for the correct temporal pacing of puberty, as demonstrated by animal models and the absence of pubertal development in patients with HH (5, 78, 79). Our tissue expression studies localized *IGSF10* mRNA expression to a spatial and temporal window when GnRH neurons are migrating through the nasal mesenchyme (E11.5-17.5 in mice) to the border with the telencephalon. Further evidence for a role of IGSF10 in the regulation of GnRH neuronal migration was gained from our *in vitro* demonstration of reduced migration of immortalized GnRH neurons into a cell milieu with strongly reduced *Igsf10* expression. These results were additionally validated by the use of a transgenic zebrafish model, where *Igsf10* knockdown via injected MOs in zebrafish embryos resulted in perturbed migration and failed neurite extension of GnRH3 neurons towards the hypothalamus.

We hypothesize that immature GnRH neurons may respond to IGSF10 signalling at the earliest stages of their migration, on exiting the VNO. Pathogenic *IGSF10* mutations, such as those identified here, will disrupt IGSF10 signalling, potentially resulting in reduced numbers or mis-timed arrival of GnRH neurons at the hypothalamus; the latter leading to a functional defect in the formation of the GnRH neuroendocrine network. With this impaired GnRH system there would follow an increased 'threshold' for the onset of puberty, with an ensuing delay in pubertal timing. This relationship has also been demonstrated in adult Reeler mice, which have significantly fewer GnRH neurons in the hypothalamus and display a phenotype of delayed pubertal maturation (59). Only the generation of specific *IGSF10* knockout mice will definitively demonstrate the number, and timing of, GnRH neurons to arrive at the hypothalamus in the absence of IGSF10 signalling.

GnRH neurons are known to have receptors for at least 20 neurotransmitters. Migratory GnRH neurons receive a plethora of guidance and movement-inducing messages during this journey, which are likely to be distinct depending on the stage of their migration (60). Signals may act directly or indirectly through the extension of olfactory axons, as disruption of the nerve tract 'scaffolds' themselves can disrupt GnRH migration (61). Gradients of chemokines may be particularly important for promoting movement of GnRH neurons (60). The pattern of *Igsf10* expression in the nasal mesenchyme shows a ventral to dorsal gradient between the VNO and olfactory bulbs, similar to a known axon-guidance gene, *SDF1* (or *Cxcl12*). *Cxcl12* and its receptor *Cxcr4* are necessary for the guidance of GnRH neurons towards the forebrain in both mouse (267) and zebrafish (268) models. In the absence of intact Cxcl12/Cxcr4 signalling, GnRH neurons fail to properly migrate out of the nasal mesenchyme during embryonic development.

The discovery of the role of *IGSF10* contributes an additional component to the highly complex system of secreted molecules and chemotactic gradients directing GnRH neuronal migration in the nasal region (55). The specific receptor(s) for secreted IGSF10 protein, on the GnRH neurons or elsewhere, remains to be determined. *IGSF10*, like *GnRH*, is also expressed in lung and other tissues, where its function is as yet unknown. A degree of redundancy with or compensation by other chemokines, such as has been postulated for the GnRH neuron migratory cue semaphorin-4D, as well as the action of potential 'protective factors', is feasible (55).

Furthermore, mutations in *IGSF10* may contribute to the phenotype of other forms of secondary hypogonadism. Mutations that perturb GnRH neuronal migration, including in *KAL1* and *PROKR2*, are already known to cause both HH and HA (181). Thus, mutations in

genes which lead to late arrival or reduced numbers of GnRH neurons to the hypothalamus, such as *KAL1*, *PROKR2* and now potentially *IGSF10*, can compromise the function of the GnRH network with variable clinical manifestations from HA to HH. Moreover, the same authors found that 25% of patients with HA in their study had DP. DP is commonly found in relatives of patients with HH and KS. These findings imply that disruption to the GnRH network may result in a spectrum of phenotypes from DP, through HA to complete GnRH deficiency. An increasing burden of mutations may produce a more severe outcome, with perhaps one mutation leading to DP or HA, whilst 2 or even more mutations may be required to lead to a phenotype of HH or KS. The identification of loss-of-function mutations in *IGSF10* in two out of 14 individuals sequenced with HA and HA-equivalent adds weight to this suggestion, although sequencing of *IGSF10* in a larger cohort of HA patients would help to confirm this assertion. However, although rare, predicted pathogenic variants in *IGSF10* were enriched in our HH cohort as compared to the general population, the presence of such variants in the public databases suggests that these variants alone are not sufficient to cause complete idiopathic HH or KS, again in keeping with the above hypothesis.

In conclusion, these findings strongly support the contention that mutations in *IGSF10* cause DP in humans, through dysregulation of GnRH neuronal migration during embryonic development. Moreover, such mutations may also underlie susceptibility to the phenotype of HA. Overall, this represents a new causal mechanism for self-limited DP, and reveals a shared pathophysiology between DP and other forms of functional hypogonadism.

Chapter 4

Results 2: A Mutation in *HS6ST1* Causes Self-Limited Delayed Puberty; Evidence for Overlap Between the Genetic Basis of Hypogonadotropic Hypogonadism and Delayed Puberty

4.1 Rare variant in *HS6ST1* identified following whole exome sequencing

As outlined in chapter 3, whole exome sequencing performed in the 18 most extensive families from our cohort (111 individuals: a total of 76 individuals with DP, male=53, female=23; and 35 controls, male=13, female=22), identified 2,474,145 variants after quality control (Fig. 4.1.1). This whole exome data was filtered through our in-house bioinformatics pipeline as described previously, to identify rare, predicted deleterious mutations, segregating with trait in an autosomal dominant inheritance pattern in multiple families and with potential biological relevance.

Of the 28 top candidate genes taken forward for targeted re-sequencing in a further 42 families (178 individuals with DP and 110 controls), only candidate gene, *HS6ST1* (ENSG00000136720, gene identification number 9394) had been identified through the use of the 'HH genes' filter (Appendix 1), as a gene already implicated in the pathogenesis of HH. Interestingly, prior to filtering for segregation within cohort families, 26 variants in 15 genes from this 'HH genes' list had been identified, highlighting the value of familial data in identification of causal variants in a condition such as DP (Table 4.1.1). Following whole gene rare variant burden testing with a multiple comparison adjustment (238) (Table 3.1.1), *HS6ST1* was retained as a candidate of interest with an adjusted p value of <0.025.

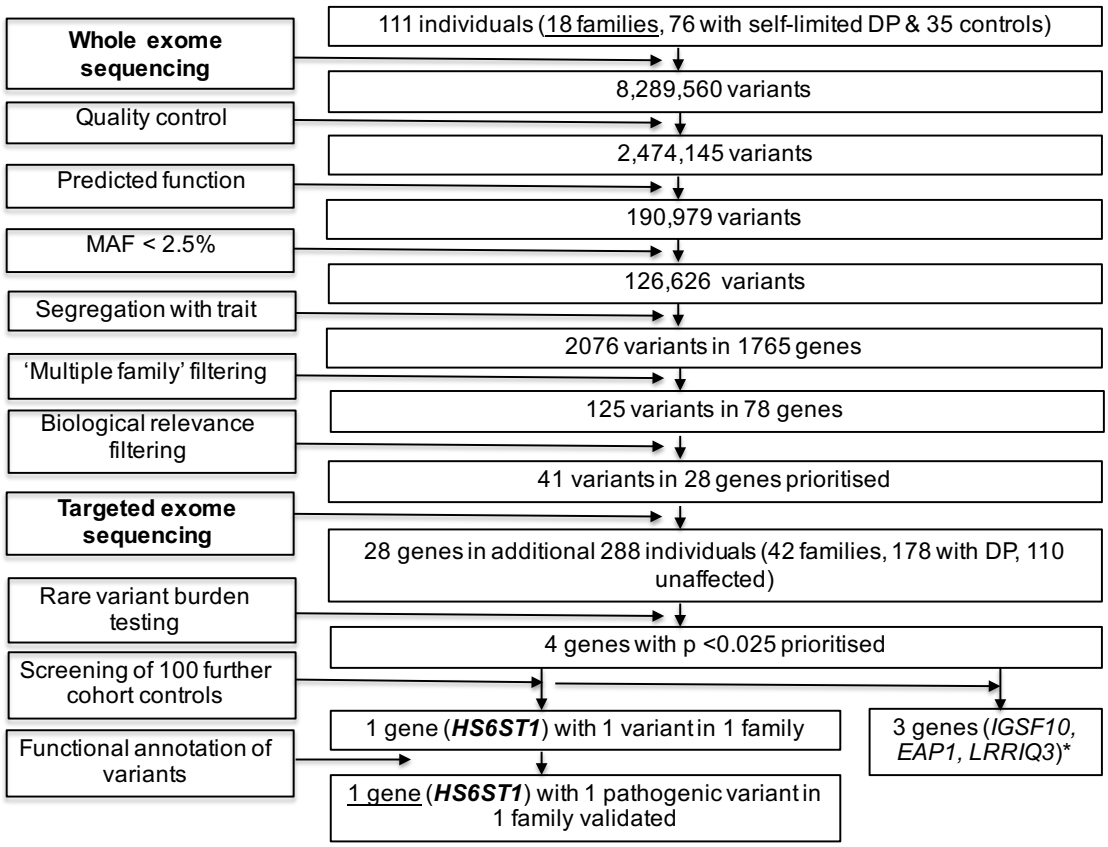


Figure 4.1.1 – Flowchart of WES (whole exome sequencing) filtering strategy to identify *HS6ST1*.

Whole exome sequencing was initially performed on DNA extracted from peripheral blood leukocytes of 111 individuals from the 18 most extensive families from our cohort (76 with DP and 35 controls), with exome capture on a Nimblegen V2 or Agilent V5 platform and sequencing on the Illumina Hiseq 2000. The exome sequences were aligned to the UCSC hg19 reference genome. Picard tools and the genome analysis toolkit were used to mark PCR duplicates, realign around indels, recalibrate quality scores and call variants. Variants were then analysed further and filtered for potential causal variants using filters for quality control, predicted functional annotation, minor allele frequency (MAF), segregation with trait, variants in multiple families and biological relevance (See methods and supplementary table 1 for further information on filtering criteria). Targeted exome sequencing using a Fluidigm array of 28 candidate genes identified post-filtering was then performed in a further 42 families from the same cohort (288 individuals, 178 with DP and 110 controls). Variants post targeted re-sequencing were filtered using the same criteria as the whole exome sequencing data. Rare variant burden testing was performed for all genes selected for targeted re-sequencing, with a multiple comparison adjustment applied to the set of 28 p values post hoc (238). Screening of 100 further cohort controls was via conventional Sanger sequencing. Functional annotation of the variants as described elsewhere in methods. DP – delayed puberty. *data published in (269) or unpublished.

Gene	No. of rare, predicted damaging variants filtered from whole exome sequencing data	
	Prior to filtering for segregation with trait	After filtering for segregation with trait
CHD7	6	0
DAX1	1	0
FGF8	0	0
FGFR1	0	0
GNRH1	0	0
GNRHR	1	0
HESX1	0	0
HS6ST1	3	1
KAL1	0	0
KISS1	2	0
KISS1R	3	0
LEP	0	0
LEPR	1	0
LHX3	2	0
LHX4	1	0
NELF	0	0
PCSK1	1	0
PROK2	0	0
PROKR2	0	0
PROP1	1	0
SEMA3A	1	0
SF1	1	0
TAC3	0	0
TAC3R	1	0
WDR11	1	0

Table 4.1.1 – HH gene filtering

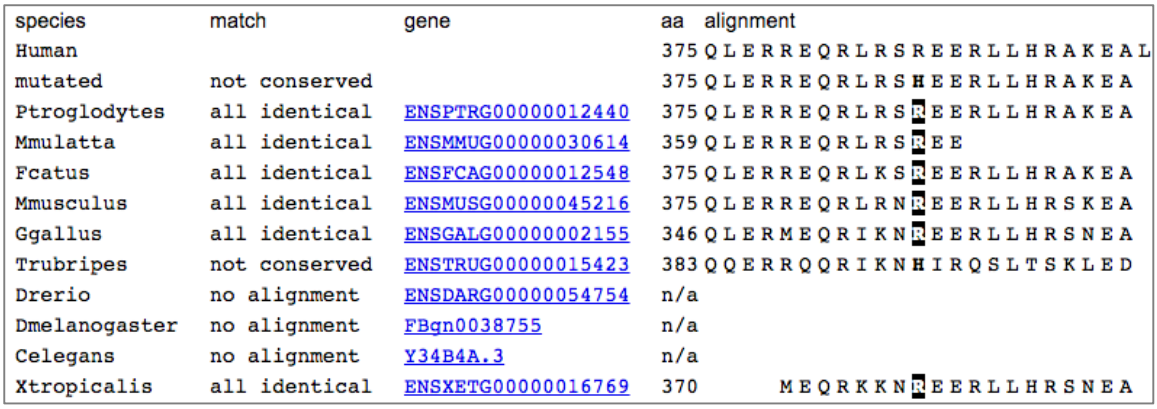
25 ‘HH genes’, mutations in which have been identified as causal in patients with GnRH deficiency, used for filtering strategy post whole exome sequencing. Details given of number of variants in these genes identified in 18 families with DP after whole exome sequencing, prior to and after filtering for segregation with trait within each family.

One rare, predicted damaging variant segregating with trait was identified in *HS6ST1* (NM_004807.2: c.1124G>A (rs182882999) p.Arg375His) in one family through whole exome sequencing. After targeted exome sequencing, two additional potentially pathogenic variants in *HS6ST1* (NM_004807.2: c199A>T (rs202247387) p.Lys67* and NM_004807.2: c585G>A p.Trp195*) were identified in four further affected individuals. However, neither variant was validated by Sanger sequencing (Fig. 4.1.1).

The one rare heterozygous missense variant that remained post filtering and verification, p.Arg375His, is predicted to be deleterious by 4/5 prediction tools (Table 4.1.2). The variant lies within a highly conserved coiled-coil domain affecting an amino acid that is highly conserved among homologues, as revealed by PhyloP and GERP score, and multiple sequence alignment (Fig. 4.1.2 and Table 4.1.2).

AA Change	**dbSNP137**	**PhyloP (254)**	**SIFT (237)**	**Polyphen 2 (236)**	**LRT (255)**	**Mutation Taster (256)**	**FATHMM (257)**	**GERP++ (258)**
p.Arg375His	rs182882999	C	D	B	D	D	D	4.3

Table 4.1.2 – Prediction of variants according to web-based prediction software programs and conservation across species. C – conserved, D – deleterious, disease causing or damaging, B - benign



species	match	gene	aa	alignment
Human			375	Q L E R R E Q R L R S R E E R L L H R A K E A L
mutated	not conserved		375	Q L E R R E Q R L R S **H** E E R L L H R A K E A
Ptroglodytes	all identical	ENSPTRG00000012440	375	Q L E R R E Q R L R S R E E R L L H R A K E A
Mmulatta	all identical	ENSMMUG00000030614	359	Q L E R R E Q R L R S R E E
Fcatus	all identical	ENSFCAG00000012548	375	Q L E R R E Q R L K S R E E R L L H R A K E A
Mmusculus	all identical	ENSMUSG00000045216	375	Q L E R R E Q R L R N R E E R L L H R S K E A
Ggallus	all identical	ENSGALG00000002155	346	Q L E R M E Q R I K N R E E R L L H R S N E A
Trubripes	not conserved	ENSTRUG00000015423	383	Q Q E R R Q Q R I K N **H** I R Q S L T S K L E D
Drerio	no alignment	ENSDARG00000054754	n/a	
Dmelanogaster	no alignment	FBgn0038755	n/a	
Celegans	no alignment	Y34B4A.3	n/a	
Xtropicalis	all identical	ENSXETG00000016769	370	M E Q R K K N R E E R L L H R S N E A

Figure 4.1.2 Multiple sequence alignment (m.s.a.) for the p.Arg375 (R375) residue.

M.s.a. was generated using MutationTaster (256). The p.Arg375 residue is highly conserved amongst different species, GERP score 4.3.

Although this variant was present at low MAF in some public databases, it was highly enriched in our cohort (Table 4.1.3). Its presence in databases is not unexpected as up to 5% of the general population will have abnormal pubertal timing, either early or delayed.

Nucleotide change	Amino acid change	Exon	MAF from DP patients (%) (n=254)	MAF from cohort controls (%) (n=180)	MAF from ExAC database (%) Finnish/European/All
c.1124G>A	p.Arg375His	2	2.4	0	0.03/0.003/0.03

Table 4.1.3 – Minor allele frequency in study population and control cohorts.

Minor allele frequency data for the control population was retrieved from the ExAC Browser (Exome Aggregation Consortium (ExAC), Cambridge, MA: http://exac.broadinstitute.org, accessed May 2015).

4.2 Rare variant in *HS6ST1* segregates with trait in one family with classical self-limited DP

The family pedigree is shown in figure 4.2.1. Clinical details of this family revealed them to have typical features of self-limited DP. The proband case was first investigated for growth delay at 12.8yrs, at which time his bone age was 11yrs. Examination at this stage found him to be prepubertal with testis volumes of 2ml bilaterally and no pubic hair development. His biochemistry at that time showed a typical picture of HH. This was followed by spontaneous onset of puberty around 14.1 yrs of age and subsequent development over 2 yrs to normal adult testosterone levels and testicular volume. His sister's age at menarche was 15yr; both had normal birth weight and birth length. Their father and paternal uncle and aunt all had DP with delayed linear growth. All cases had self-reported normal olfaction.

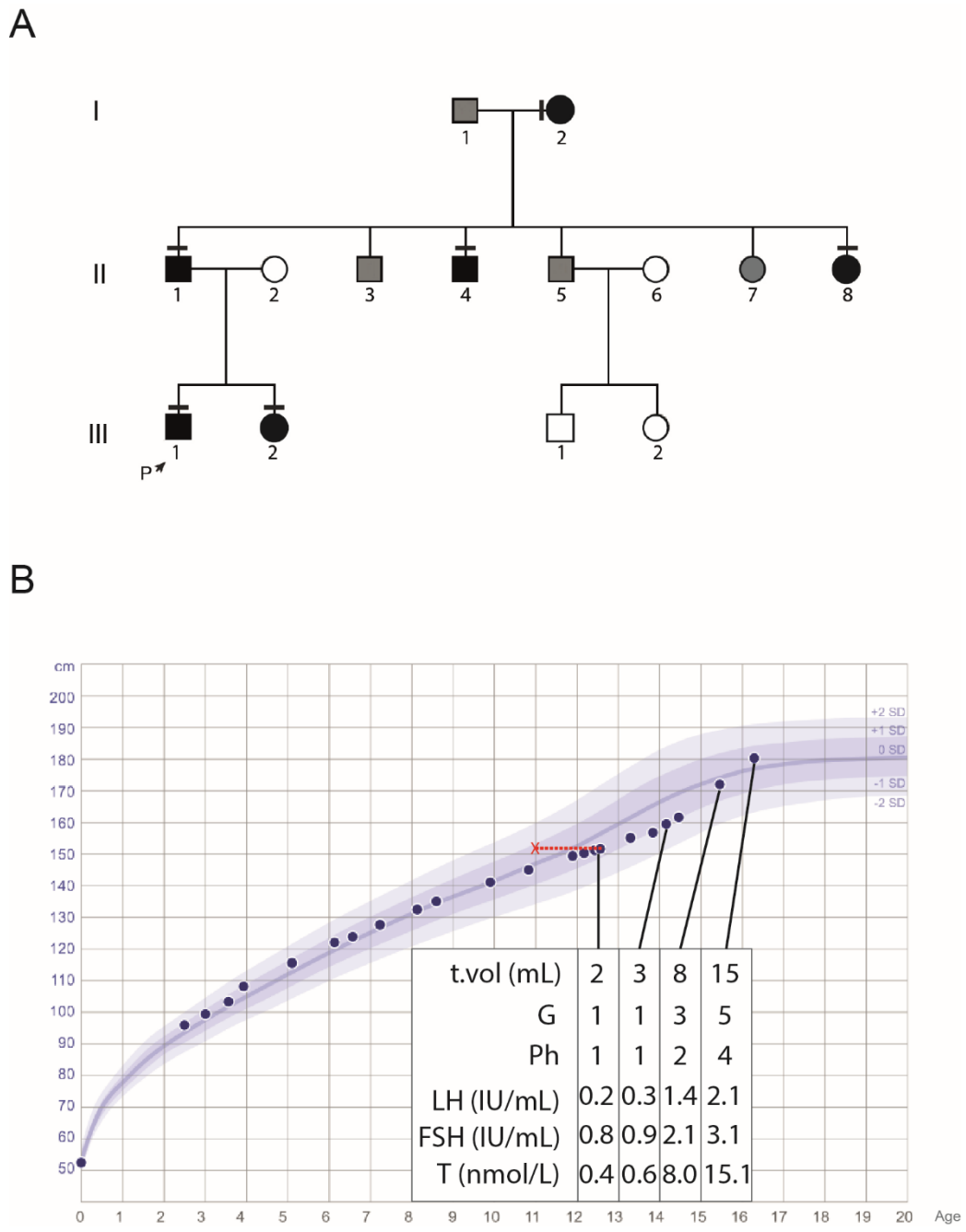


Figure 5.2.1 – Pedigree of the family with *HS6ST1* p.Arg375His mutation

A: Squares indicate male family members, circles female family members. Black symbols represent clinically affected, grey represent unknown phenotype, clear symbols represent unaffected individuals. The arrow with 'P' indicates the proband in each family. A horizontal black line above an individual's symbol indicates they are heterozygous for that mutation as confirmed by either whole exome sequencing or Fluidigm array, and verified by Sanger sequencing.

B: Height chart for the proband showing reduction in growth velocity from 12 yrs of age with associated delayed bone age shown in red. The subject was prepubertal until 14.1 yrs of age. This was followed by spontaneous onset of puberty and subsequent development over 2 yrs to normal adult testosterone levels and testicular volume. T vol – testicular volume, G – Tanner genital stage, Ph – Tanner pubic hair stage, LH – luteinising hormone, FSH – follicular stimulating hormone, T – testosterone

4.3 HS6ST1 mutant protein displays reduced sulphotransferase activity *in vitro*

In vitro analysis demonstrated reduced sulphotransferase activity of the p.Arg375His mutant protein (Fig. 4.3.1), with mean +/- SD activity as compared to wild-type protein of 39.8% +/- 8.2 (n=3, p<0.001). This reduction in enzymatic function is within the range seen with other pathogenic mutations in *HS6ST1* (77). However, the relative activity of the p.Arg375His was not as low as the p.Met404Val activity, suggesting that the impact of this mutation is not as severe as some of the most damaging mutations seen in patients with HH (77).

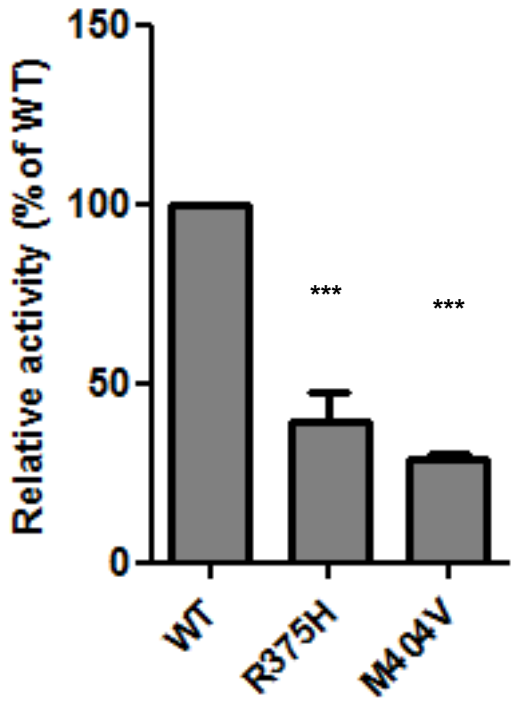


Figure 4.3.1 – The p.Arg375His mutation reduces HS6ST1 sulphotransferase activity *in vitro*. Relative specific activity of recombinant WT or mutant HS6ST1 protein is shown. All experiments were done in triplicate using equal amounts of protein. *** P<0.001

4.4 *Hs6st1* is expressed in both the nasal placode of developing embryos and within the peri-pubertal hypothalamus

To gain insight into the role that *HS6ST1* might be playing in the timing of puberty, we examined tissue expression by RT-PCR and *in situ* hybridisation. RT-PCR analysis indicated expression of *Hs6st1* in both immortalised immature migratory GnRH neurons (GN11 cells) and immortalised mature non-migratory (GT1-7 cells) (261, 262), as well as in mouse hypothalamus (Fig. 4.4.1).

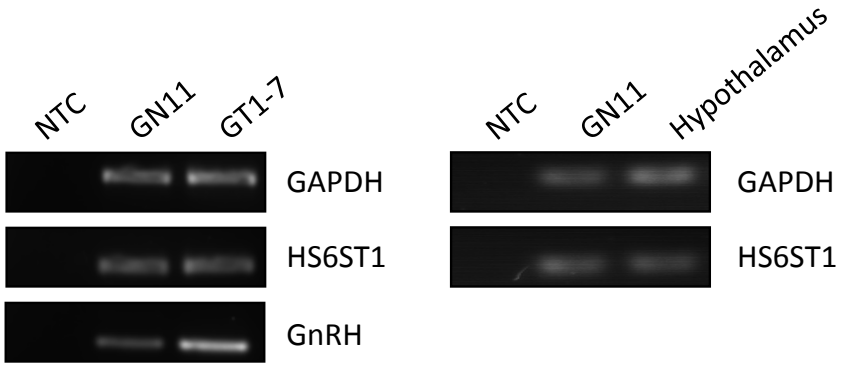


Figure 4.4.1 – Tissue expression of *Hs6st1* by RT-PCR. RT-PCR analysis showed expression of *Hs6st1* within both immortalised immature migratory GnRH neurons (GN11 cells) and immortalised mature non-migratory (GT1-7 cells), as well as in mouse hypothalamus. NTC – no template control

GnRH neurons are believed to originate from the nasal placode, and are first visualized within the vomeronasal organ (VNO) at around E10.5. *Hs6st1* mRNA expression was detected in the nasal placode from E11 through to E14.5, with strong expression within the VNO itself, as well as in the olfactory epithelium (OE) (Data from E14.5 shown in Fig. 4.4.2, panel A). GnRH neurons could be visualised migrating through nasal mesenchyme at E14.5

with *Hs6st1* expression along the vomeronasal nerve, along which the GnRH neurons travel towards the olfactory bulb (Fig. 4.4.2, panels C&D). Additionally, expression of *Hs6st1* mRNA was seen in peri-pubertal mice within the hypothalamus, but conclusive evidence for co-localisation with *Gnrh* expression was not found (Fig. 4.4.2, panels E&F).

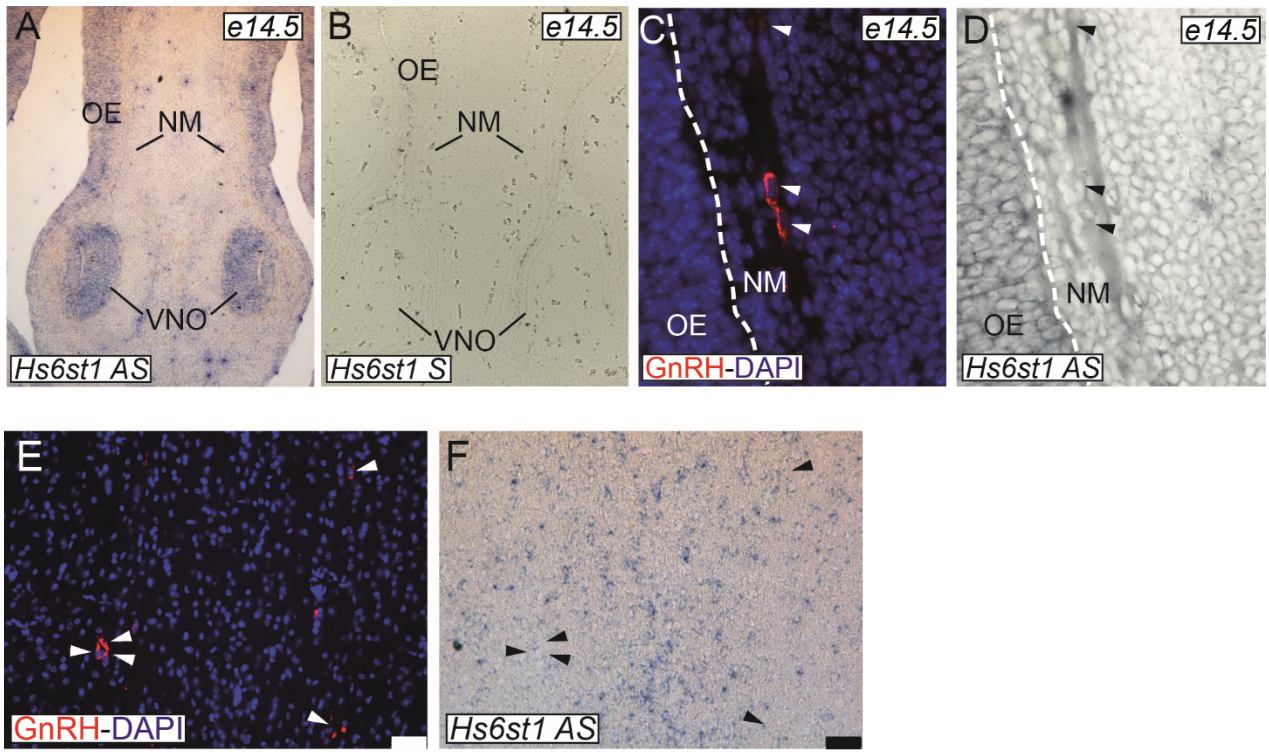


Figure 4.4.2 - Expression pattern of *Hs6st1* mRNA in mouse developing brain

A: Hs6st1 expression (in purple) was observed at E14.5 with strong expression seen in the vomeronasal organ (VNO) and olfactory epithelium (OE), with low level expression in the nasal mesenchyme (NM). B: Mouse sense probe resulted in no specific signal. C&D: GnRH neurons could be visualised migrating through nasal mesenchyme at E14.5 with *Hs6st1* expression along the vomeronasal nerve (D, dark grey), whilst surrounding tissue expressed *Hs6st1* at low levels. E&F: In peri-pubertal female mice, *Hs6st1* expression (F, in purple) was seen with the hypothalamus with some overlap of localisation with *GnRH* signal. GnRH neurons are shown in pink and with arrowheads, DAPI in blue (panels C and E). S – sense, AS - antisense.

A similar expression pattern of *HS6ST1* was detected in the nasal area of human embryos at 9 post-conceptual weeks (pcw), again with apparent low-level expression of *HS6ST1* in GnRH neurons. However, although clusters of GnRH neurons were seen within the nasal placode (Fig 4.4.3, panels A-C), the identification of human VNO, a structure which is notoriously difficult to find, was not conclusive (270).

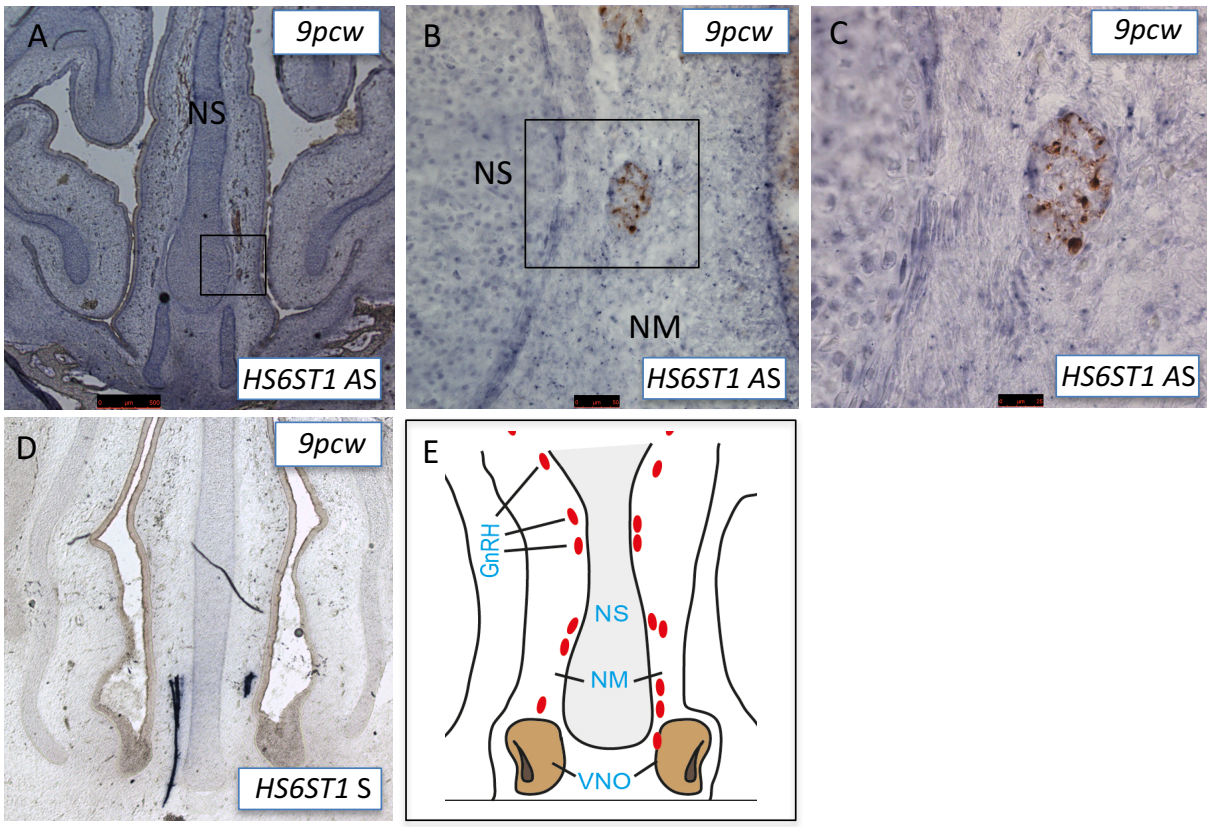


Figure 4.4.3 - Expression pattern of *Hs6st1* mRNA in human developing brain

In human 9pcw brains, *HS6ST1* expression pattern (in purple) was similar to that observed in mouse, with clusters of GnRH neurons seen on the background of low level *HS6ST1*-expressing nasal mesenchyme (NM) (A-C, frontal sections, see also schematic in E). D - sense probe resulted in no specific signal. GnRH neurons are shown in brown.

4.5 Discussion

In view of the possible overlap between the pathophysiology of DP and HH, screening DP patients for mutations in known HH genes appears a prudent strategy; however, minimal overlap between the two has been identified to date (226). Our results have identified a deleterious mutation in *HS6ST1 (Heparan sulfate 6-O-sulphotransferase 1)* as the likely causal factor in the development of self-limited DP in one pedigree from our large cohort of patients with familial DP. *HS6ST1* mutations have been previously identified in up to 2% of patients with idiopathic HH (77, 227); despite this, they have not previously been found in pedigrees with familial self-limited DP. Furthermore, our investigations provide expression data implicating a role for *HS6ST1* in the development of the GnRH neuronal network.

Extracellular matrix components such as Heparan Sulfate (HS) proteoglycans can act as regulators of axon guidance (271), and have been shown to be crucial for neural development in mice (272). HS regulate the biological activity of numerous signalling proteins via sugar-protein interactions (273), including those known to be key for the regulation of GnRH neuronal development such as Slit/Robo and FGF/FGFR (274). The specificity and functionality of FGFs are controlled by modification of sugar residues by sulfations, that are introduced by HS-modifying enzymes such as *HS6ST1*. *HS6ST1*, one of three *HS6ST* isoforms, codes for an enzyme that transfers a sulfate residue to position 6 of the GlcNAc generating $GlcNSO_3$ (275). The majority of mice deficient in *Hs6st1* (*Hs6st-1*$^{-/-}$) die at E15.5, while those which develop further exhibit aberrant placental and lung morphology and disrupted axon navigation in the optic chiasm and corpus callosum (276).

Hs6st1 is known to be required for the function of *Fgfr1* and *Anos1* (formally known as *Kal1*) *in vivo*, both of which are vital for normal HPG axis function (77). Using a *C. elegans* model,

Tornberg et al (77) provided evidence that *Hs6st1* may act to regulate neural branching in concert with *Anos1* and *Fgfr1.* Fgfr relative affinity for fibroblast growth factors is increased by HS glycosaminoglycans that function as co-receptors (277). Fgf8 protein levels were shown to be elevated in the embryonic telencephalon of *Hs6st-1*$^{-/-}$ mice (273). *Anos1* is also able to enhance *Fgfr1* signalling in a HS dependent manner and it has been proposed that *Anos1* finely tunes *Fgfr1* signalling during olfactory and GnRH neuronal development (278). However, *Hs6st1* mutations may also impact on phenotype via *Anos1* independent pathways (77).

In vitro analysis of the p.Arg375His mutant protein as compared to wild type demonstrated a reduction in sulphotransferase activity within the range of previously published data (77). However, some known mutations in *HS6ST1* leading to HH, e.g. p.Met404Val, result in a more dramatic reduction in enzymatic activity. This may indicate a spectrum of HS6ST1 inactivity related to human phenotype, with partial function in the heterozygous state resulting in a less severe phenotype of self-limited DP, whilst more deleterious mutations or a 'second-hit' in a separate gene produce the phenotype of HH.

Hs6st1 is expressed both at sites of GnRH neuronal development in the embryonic nasal placode, and within the peri-pubertal hypothalamus in mice. High expression of *Hs6st1* in the VNO suggests it may have importance for GnRH neuronal early developmental or migratory events, although expression in the hypothalamus does not rule out a potential role of Hs6st1 in GnRH neuronal function once these neurons have reached their final position. These findings are in keeping with published literature that Hs6st1, through its interactions with both Anos1 and Fgfr1, may have role at more than one time-point in determining GnRH neuronal biology. FGF signalling is required for olfactory placode

induction and differentiation, as well as GnRH neuronal fate specification and postnatal maintenance (279). It remains unclear whether Hs6st1 functions through its activity within GnRH neurons or via expression in adjacent cells.

Hs6St1 is also thought to be important for glial cell positioning: *Hs6st1*$^{-/-}$ mice show ectopic clumping of glial cells in the midline/ area of the developing corpus callosum (274), also seen in conditional FGFR1-/- mutant embryos. (280) Glial cell inputs facilitate GnRH secretion via growth-factor derived signalling (80). Thus, glial cell dysfunction may be a further mechanistic link between mutations in HS6ST1 and the phenotype of DP.

In summary, we have identified a new pathogenic mutation in *HS6ST1* as the likely cause of DP in a pedigree from our extensive cohort. Interestingly, no other pathogenic mutations in genes known to cause HH were identified in our cohort in this study, suggesting that the genetic background of HH and DP may be largely different, or shared by as yet undiscovered genes. Our findings do contrast somewhat with previous publications of up to 14% of DP probands carrying potentially pathogenic variants in HH genes (226). However, previous results have not been adjusted for segregation with the DP trait within families nor have the variants been tested for pathogenicity, and thus may be an over-estimation. However, the results from our study deserves validation in further pedigrees with familial DP. Overall, our findings of a deleterious mutation in *HS6ST1* in a family with self-limited DP provides further evidence that perturbations in the development or function of GnRH neurons, perhaps relativity subtle in nature, may lead to 'simple' DP, whilst more deleterious alterations will result in more severe phenotypes such as HH.

Chapter 5

Results 3: Mutations in the gene *FTO* in Self-Limited Delayed Puberty; Evidence for Overlap Between the Genetic Basis of the Timing of Puberty in the General Population and Delayed Puberty

5.1 Mutations in the *FTO* gene identified following exome sequencing

Whole exome sequencing data from 18 informative families from our cohort with self-limited DP was filtered through our in-house bioinformatics pipeline as described previously, with the aim of identifying rare, predicted deleterious mutations, segregating with trait in an autosomal dominant inheritance pattern in multiple families and with potential biological relevance. In order to address the question of to what extent variants in the same genes cause both self-limited DP and lead to variation in the timing of puberty in the general population, the filtered variants were then searched for genes that lie within linkage disequilibrium with loci identified as associated with the timing of puberty by GWAS (Fig. 5.1.1).

Following whole exome sequencing, 16 rare and potentially pathogenic variants in 11 genes that are in linkage disequilibrium with loci identified as associated with the timing of puberty by GWAS were identified. After targeted exome sequencing and whole gene rare variant burden testing, 4 of these 11 genes with $p < 0.02$ (adjusted $p < 0.1$ post correction) were prioritised (Table 5.1.1 and Fig. 5.1.2). 2 genes, *SEC24A* and *ZNF560* were furthermore excluded as the variants identified within them were also found in multiple control subjects. The 2 remaining genes, *FTO* and *LGR4*, were retained for further analysis. *LGR4* will discussed in a later section (see Chapter 7).

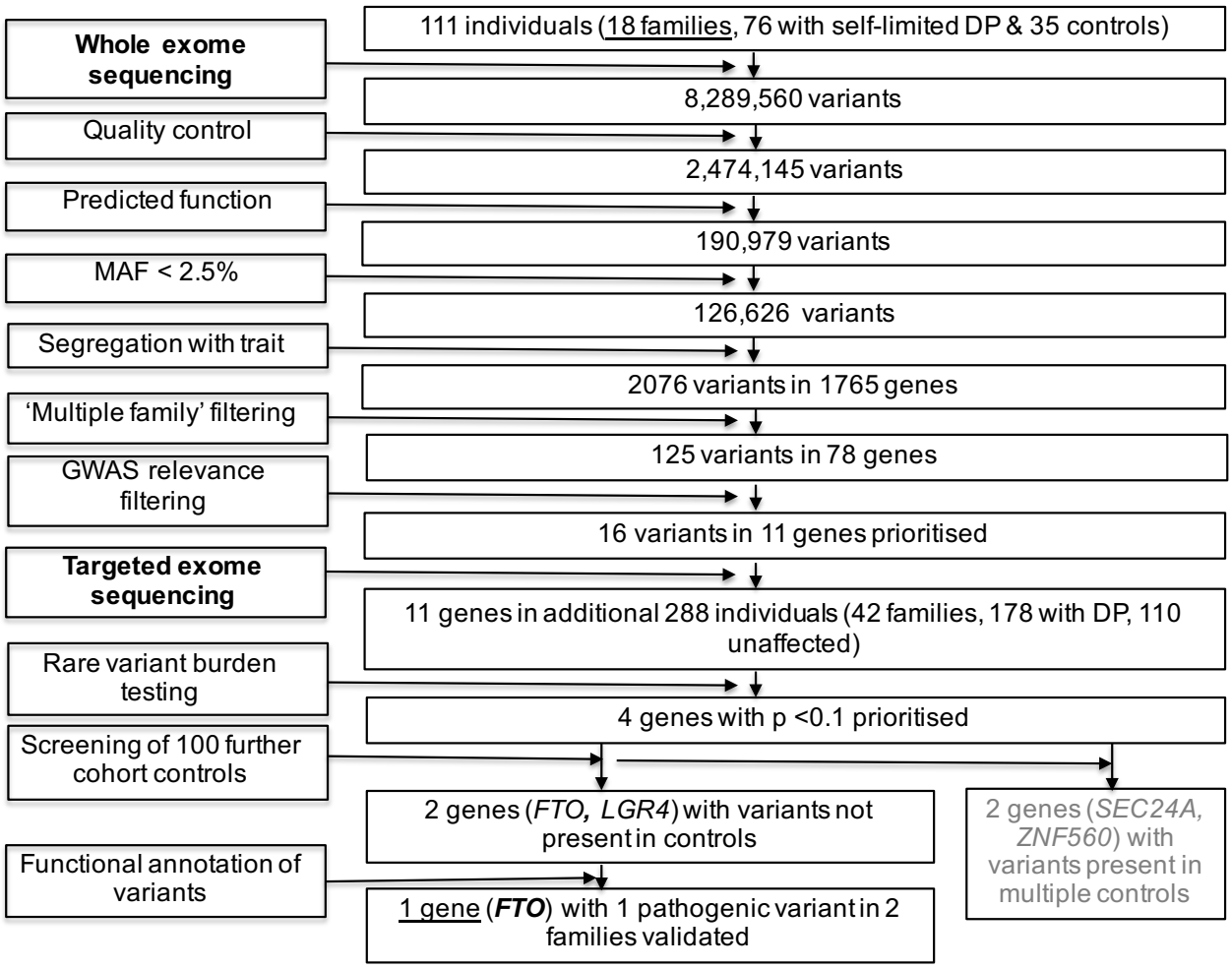


Figure 5.1.1 – Flowchart of whole exome sequencing filtering strategy to identify *FTO*.

Whole exome sequencing was initially performed on DNA extracted from peripheral blood leukocytes of 111 individuals from the 18 most extensive families from our cohort (76 with DP and 35 controls), with exome capture on a Nimblegen V2 or Agilent V5 platform and sequencing on the Illumina Hiseq 2000. The exome sequences were aligned to the UCSC hg19 reference genome. Picard tools and the genome analysis toolkit were used to mark PCR duplicates, realign around indels, recalibrate quality scores and call variants. Variants were then analysed further and filtered for rare, potentially pathogenic variants with GWAS relevance (See methods for further information on filtering criteria). Targeted exome sequencing using a Fluidigm array of 11 candidate genes identified post-filtering was then performed in a further 42 families from the same cohort (288 individuals, 178 with DP and 110 controls). Variants post targeted re-sequencing were filtered using the same criteria as the whole exome sequencing data. Rare variant burden testing was performed for all genes selected for targeted re-sequencing, with multiple comparison adjustment applied post hoc (238). Screening of 100 further cohort controls was via conventional Sanger sequencing. Functional annotation of the variants as described in methods. DP – delayed puberty.

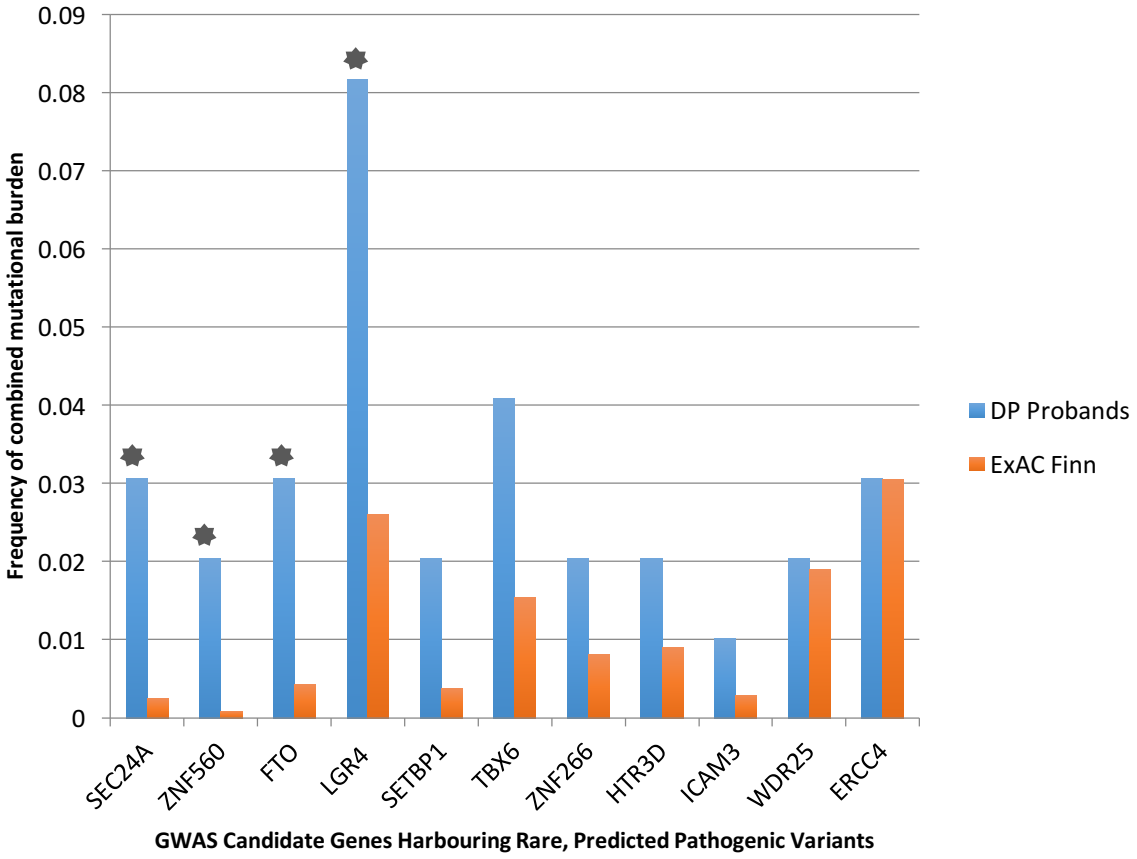


Figure 5.1.2 - Mutational burden for all GWAS AAM genes with rare predicted pathogenic variants in DP patients compared to healthy control individuals from the ExAC (Finnish) Database.

Frequency of combined mutational burden is shown for the 11 genes harbouring rare, functional predicted pathogenic variants in DP probands (n=49) vs ExAC (Finnish) healthy controls (n=3305). The four genes with a significant enrichment in cases versus controls are indicated by an asterisk (Fisher's Exact Test, $p<0.02$).

Gene	Total number of rare and predicted damaging* variant alleles in probands and in the reference population		P value	Adjusted P value**
	In DP cohort (n=49)	In ExAC Finn database (n=3305)		
SEC24A	3	17	0.0051	0.029
ZNF560	2	6	0.00823	0.038
LGR4	8	172	0.0155	0.058
FTO	3	29	0.0188	0.058
SETBP1	2	25	0.0829	0.155
TBX6	4	102	0.1204	0.211
ZNF266	2	54	0.236	0.367
HTR3D	2	60	0.3035	0.433
ICAM3	1	20	0.315	0.433
WDR25	2	126	1	1
ERCC4	3	202	1	1

Table 5.1.1 - Rare Variant Burden Testing post Targeted Exome Sequencing for GWAS genes

Minor allele frequency (MAF) data was retrieved from the ExAC Browser (Exome Aggregation Consortium (ExAC), Cambridge, MA: http://exac.broadinstitute.org, accessed October 2015). *(MAF <2.5%), predicted damaging by both SIFT (237) and Polyphen2 (236). P value calculated by Fisher's exact test. **Benjamini & Hochberg method (238)

5.2 *FTO* gene details

Of the two remaining candidate genes, one, *FTO* (ENSG00000140718, gene identification number 79068), was of particular interest as it was known to be involved with energy homeostasis and growth (198). *FTO* (Fat mass and obesity-associated protein) has been previously described in the literature and is known to act as an Fe(II) 2-OG (2-oxoglutarate) -

dependent dioxygenase to repair alkylated DNA and RNA by demethylation (252). *FTO* contributes to the regulation of energy expenditure and homeostasis, and thus to the regulation of body size and body fat accumulation. Both $Fto^{-/-}$ and $Fto^{+/-}$ mice display significantly reduced weight gain in response to a high fat diet, accumulate less white adipose tissue and have reduced serum leptin levels (281).

Two variants in *FTO* (NM_001080432.2: c.130C>G p.Leu44Val and NM_001080432.2: c.487G>A (rs145884431) p.Ala163Thr) were identified in three families from our cohort and found in one or fewer control subjects. Both variants are rare (MAF < 0.2%) heterozygous missense variants. Although both are predicted to be benign or tolerated by >2/5 prediction software tools, one variant (p.L44V) is contained within a secondary helix structure.

5.3 *In silico* analysis

In silico analysis was carried out using the solved structure of FTO (PDB identifier: 3lfm) to determine the possible pathogenicity of the identified variants. The hydrophobic residue leucine 44 is on the surface of FTO and is part of a solvent exposed alpha helix. Substitution with valine is not predicted to alter the structure FTO or its interaction with the iron molecule or DNA. Valine and leucine have similar physico-chemical properties and introduction of valine is not predicted to disrupt alpha helix formation or to produce steric clashes.

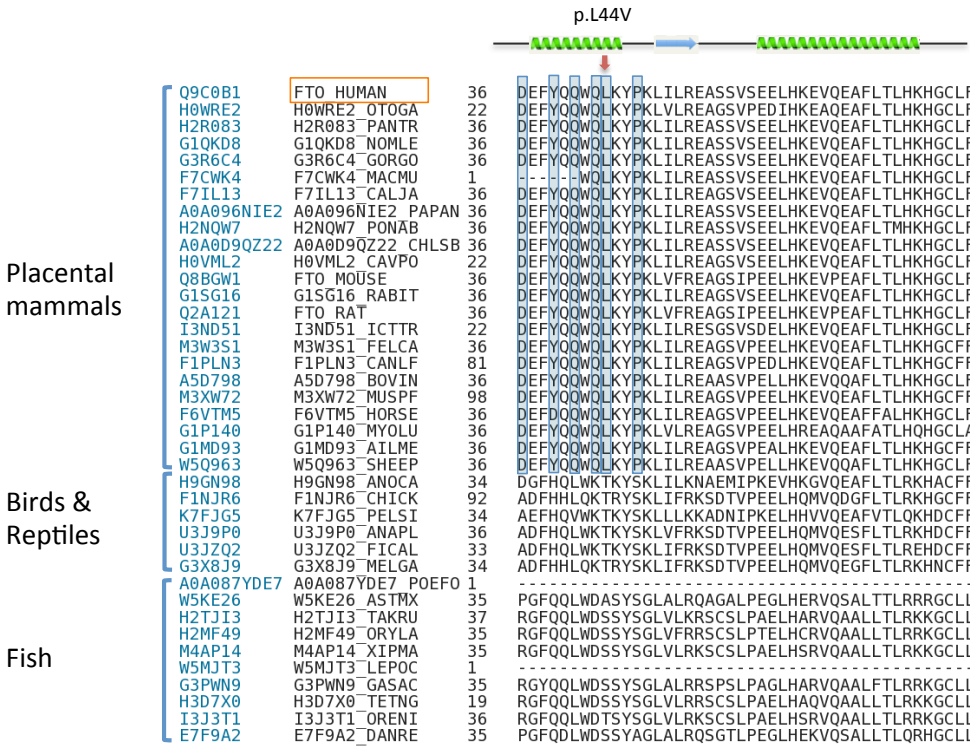


Figure 5.3.1 - Multiple sequence alignment between human FTO and its orthologues. Orthologous sequences from other species were retrieved from Ensembl. Species are classified in "Placental mammals", "Birds & Reptiles" and "Fish" according to Ensembl classification. For each species, FTO Uniprot accession number (in blue) and entry name (in black) are presented at the beginning of the row. The position of deleterious mutations on the human FTO sequence is indicated by a red arrow.

However, it is interesting to note that leucine 44 and other surrounding residues part of the same alpha helix form a motif, which is highly conserved across placental mammals but not in reptiles, birds and fish (Fig. 5.3.1). Leucine 44 is on the surface of the FTO protein (Fig. 5.3.2 and 5.3.3). Hydrophobic residues are generally in the core of the protein, away from water. When found on the surface, they may represent an interaction site between the query protein and another protein. Unfortunately, there are currently no solved structures of FTO in complex with another protein. FTO seems to have evolved a motif that is specific for mammals. This motif (residues 36-48) is a patch on FTO's surface (Fig. 5.3.4), which may

represent a mammal-specific interaction site between FTO and another protein (or DNA, if this coils around FTO), which is important for FTO function in terms of sexual development. In this case, a small change in side chain volume, such as leucine-to-valine, may produce a subtle change in protein-protein interaction and lead to a change in FTO activity *in vivo*.

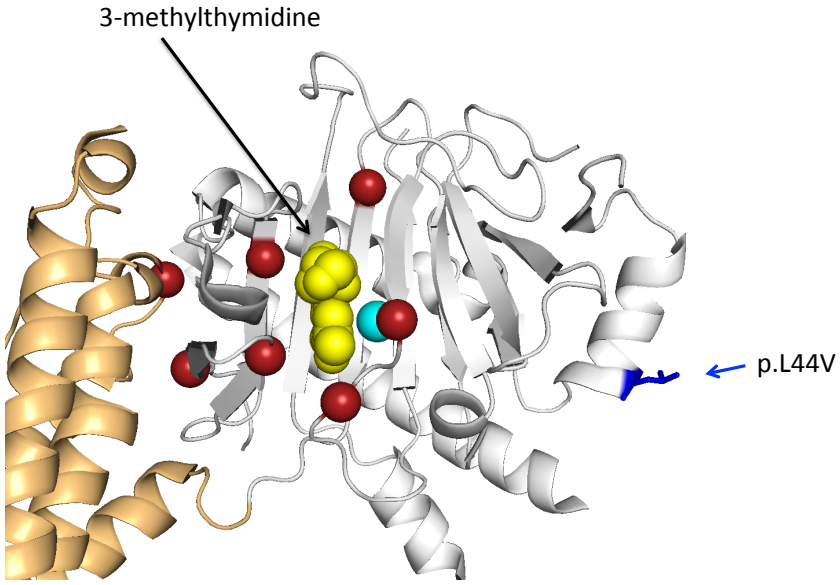


Figure 5.3.2 - The 3D structure of FTO bound to 3-methylthymidine and iron (PDB 3flm). FTO N-terminal domain is presented in grey and the C-terminal domain in orange. DNA is presented as yellow spheres, iron as a blue sphere. Deleterious mutations p.R96M or W, p. Y108A, p.F114D, p.E234P, p.C392D are presented as red spheres. p.L44V is presented in blue.

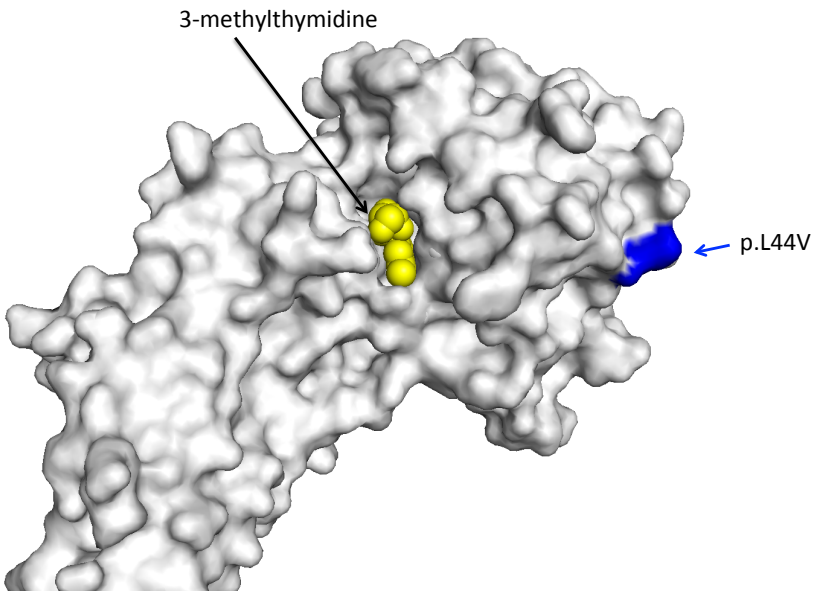


Figure 5.3.3 - Surface representation of FTO (here presented in grey) bound to 3-methylthymidine (in yellow). Leucine 44 is presented in blue.

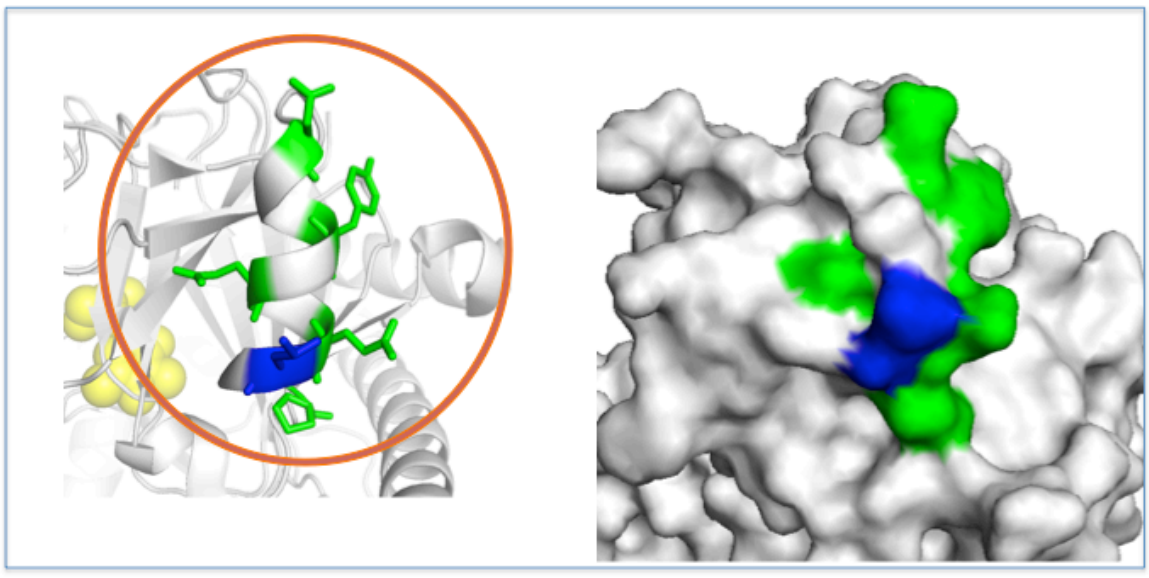


Figure 5.3.4 - Tertiary structure of FTO local to L44 residue. Residue L44 is presented in blue and other residues located on the same alpha helix and conserved across placental mammals are presented in green. FTO structure is shown as cartoon on the left and as surface on the right. 3-methylthymidine is presented as yellow spheres in the cartoon representation.

Alanine 163 is a hydrophobic, less evolutionary conserved residue, (GERP++ score 0.178, Fig. 5.3.5, panel A). Alanine 163 is at the end of the H4 alpha helix and the beginning of a long, disordered region (Fig. 5.3.5, panel B), which connects helices H4 and H5 (Fig. 5.3.5, panel C). Alanine-to-threonine substitution at this position is unlikely to affect FTO structure.

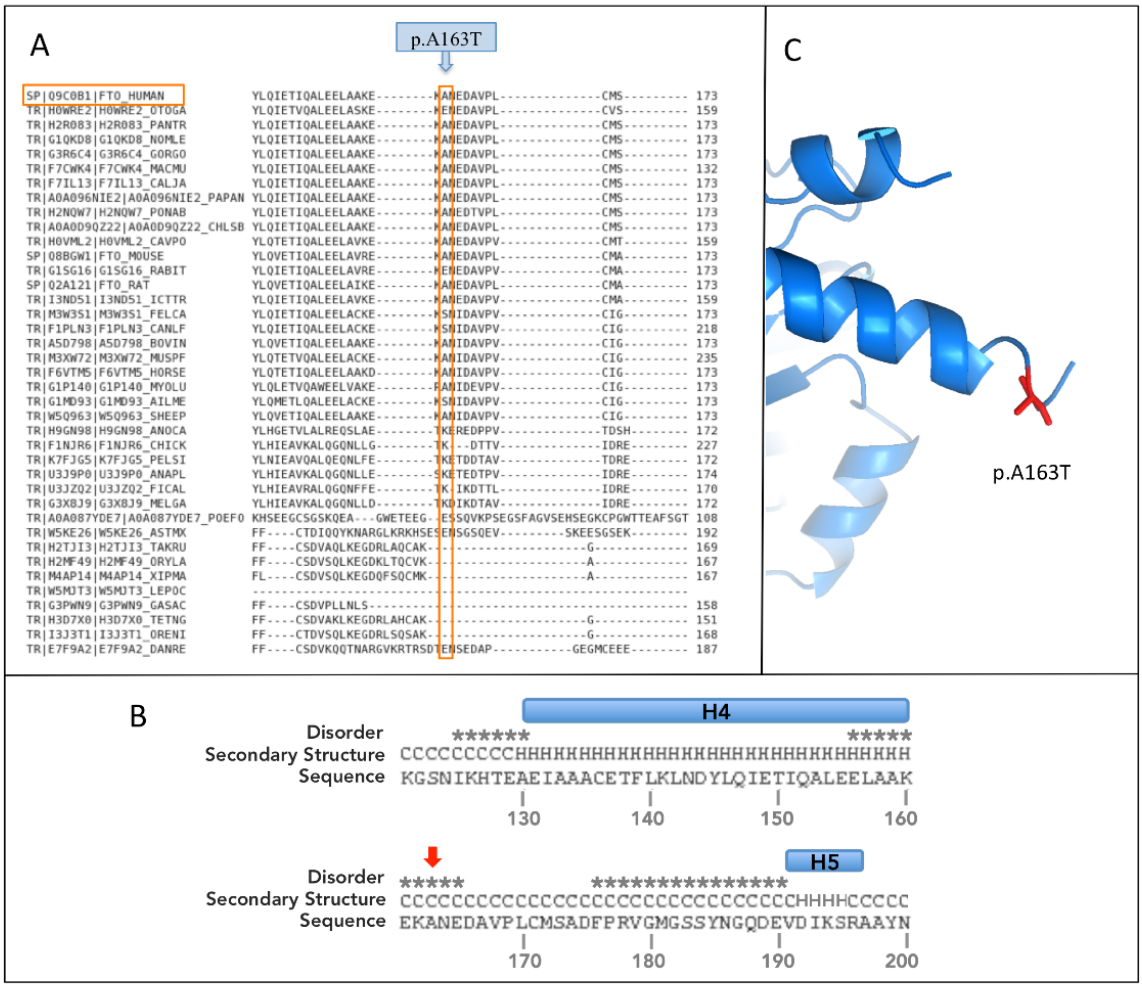


Figure 5.3.5 - p.A163T sequence and structural analysis. Panel A: a multiple sequence alignment shows the level of conservation of alanine at position 163. Alignments are listed in order of species classification by Ensembl. Panel B: the position of Ala163 (indicated by a red arrow) is presented in relation to FTO secondary structure (H, alpha helix; C, residues that are not part of an alpha helix or a beta strand). Residues predicted to be disordered are indicated with an asterisk. Panel C: FTO tertiary structure. Ala163 (presented in red) is at the end of FTO alpha helix H4 and is not predicted to disrupt FTO function. FTO structure is visualized using visualisation program (http://www.pymol.org/).

5.4 Clinical Data

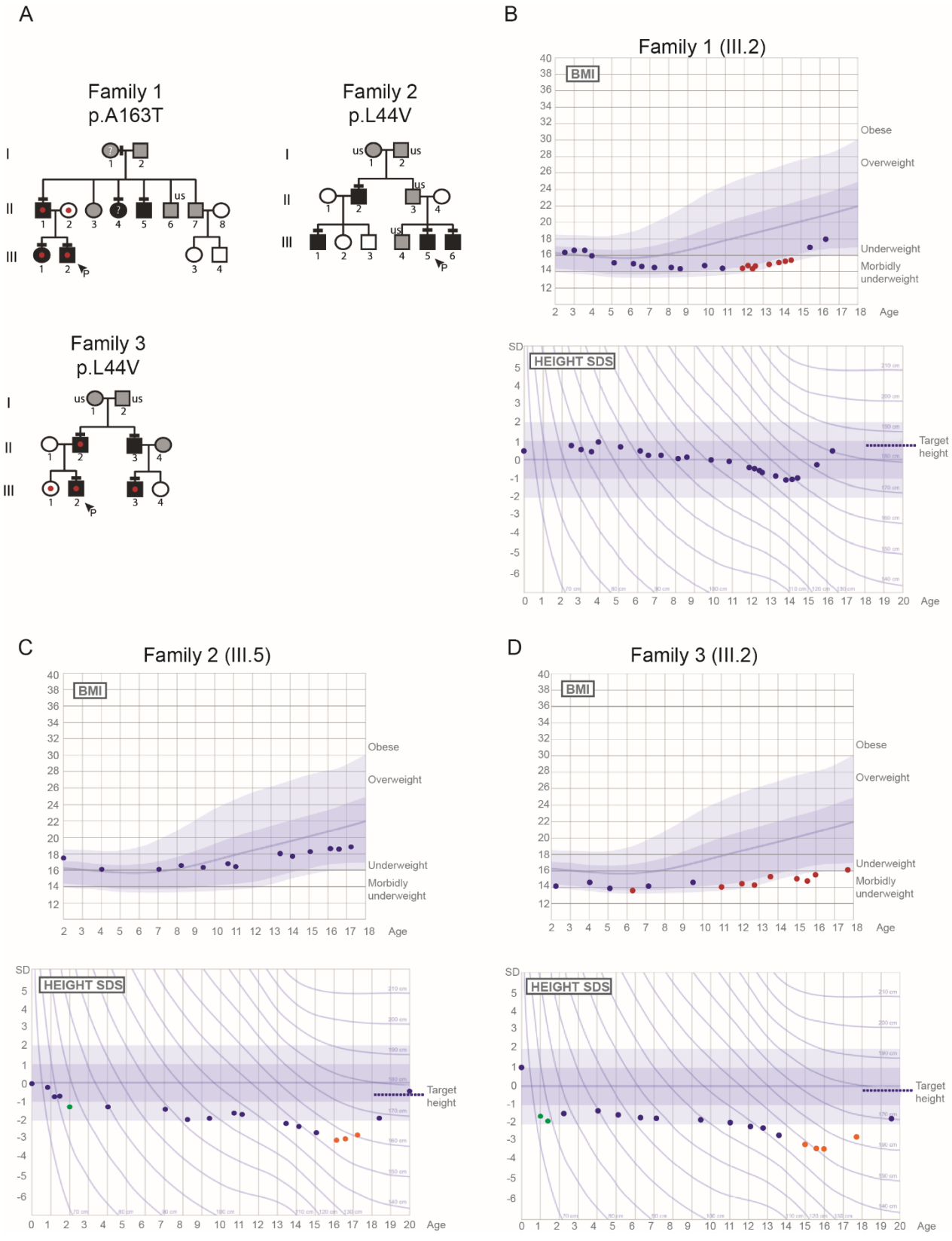


Figure 5.4.1 – Pedigrees and auxological data of the families with *FTO* variants

Panel A: Squares indicate male family members, circles female family members. Black symbols represent clinically affected, grey represent unknown phenotype, clear symbols represent unaffected individuals. The arrow with 'P' indicates the proband in each family and 'us' indicates un-sequenced due to lack of DNA from that individual. The mutation in each family is given next to the family number; a horizontal black line above an individual's symbol indicates they are heterozygous for that mutation as confirmed by either whole exome sequencing or Fluidigm array, and verified by Sanger sequencing. A red dot indicates the individual was underweight (thinness grade 2 or more significant) and '?' indicates that BMI information for that individual is not available.

Panels B-D: BMI and height standard deviation score (SDS) charts for the probands of each of the three pedigrees (family 1.III.2, family 2.III.5 and family 3.III.2). Underweight values are shown in red, green dots indicate a significant deflection from previous height measurements and orange dots indicate significant deflection from target height. Normal values, based on data from >70,000 healthy Finnish children, have been previously published [(239)].

The family identified with the p.A163T variant (family 1) and both of the families with the p.L44V variant (families 2 and 3) displayed the typical autosomal inheritance pattern of the DP trait, with perfect segregation of the variant with the trait (Fig. 5.4.1, panel A). Affected individuals from family 1 with the p.Ala163Thr variant and from family 3 with the p.Leu44Val variant were particularly underweight in childhood, with the two probands from these families (individuals 1.III.2 and 3.III.2) falling into the thinness grade 2 category (282) during the pre-pubertal period. Although there was some variability in this phenotype amongst those carrying the variants, all family members carrying *FTO* variants had ISO-BMI values in the lower range (<23) (Fig. 5.4.1 and Table 5.4.1). In addition, both of the probands from families 2 and 3 who carry the p.Leu44Val displayed faltering growth in early childhood. Both had significant defection from target height in the 2 years following birth, as well as height significantly below target height in later adolescence associated with delayed pubertal growth (Fig. 5.4.1, panels C and D) (239).

Case	Sex	Amino acid alteration	Height SDS at age 4 yrs	Height SDS at age 8/9 yrs	Height SDS at age 18 yrs	ISO-BMI at 18 yrs
1.II.1	M	p.Ala163Thr	-	1.1	1.7	16.9
1.III.2 (P)	M	p.Ala163Thr	1.1	0.5	1.1	17.1
1.III.1	F	p.Ala163Thr	0.9	1.0	1.1	17.3
1.II.5	M	p.Ala163Thr	-1.0	-1.0	-0.4	-
2.III.5 (P)	M	p.Leu44Val	-0.9	-1.4	-1.5	18.8
2.III.6	M	p.Leu44Val	-1.1	-1.3	-	-
2.II.2	M	p.Leu44Val	-	-0.8	-0.8	20.5
2.III.1	M	p.Leu44Val	0	-1.4	-	-
3.II.2	M	p.Leu44Val	-	-1.0	-0.9	18.6
3.III.2 (P)	M	p.Leu44Val	-0.9	-1.1	-1.3	18.7
3.II.3	M	p.Leu44Val		-0.4	-0.1	22.7
3.III.3	M	p.Leu44Val	-0.1	0.2	0.5	17.8

Table 5.4.1 – Clinical data of probands with *FTO* variants

Height is expressed in s.d. score (SDS) for national reference data for Finland at 4 years of age and at either 8 years for girls or 9 years for boys. Normal limits: delta HSDS <1.21, distance to target height at 4 yrs <1.76, distance to target height at 8/9 yrs <1.72 (239)

5.5 FTO mutant protein *in vitro* analysis: FTO p.Leu44Val mutant protein displays mildly reduced demethylase activity *in vitro*

In collaboration with Giles Yeo's group (Metabolic Research Laboratories and MRC Metabolic Diseases Unit Cambridge) we attempted functional characterisation of our

identified mutant FTO proteins (p.L44V and p.A163T), by assessment of the demethylation activity as compared to wildtype (WT) protein. An RNase-cleavage assay, previously verified by Dr Yeo's group, was used to measure the demethylation activity of FTO on 3-meU. Although kinetic activity of the mutant protein p.Ala163Thr did not vary from WT, mutant protein p.Leu44Val showed an approximately 20% lower kinetic activity than WT activity (Fig. 5.5.1).

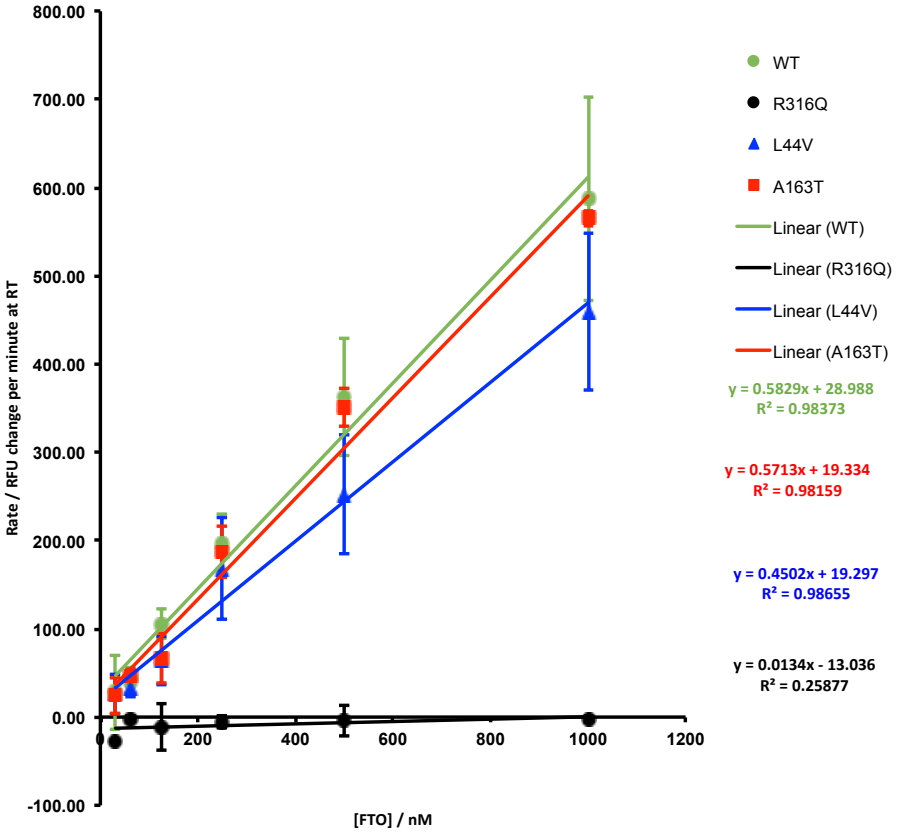


Figure 5.5.1 – Demethylation assay assessing kinetic activity of mutant versus wild type FTO proteins. FTO activity is proportional to the concentration present in the reaction. The R316Q mutant is enzymatically dead across all concentrations tested. The A163T and L44V mutants showed demethylase activity towards methylated-uridine in a dose-dependent manner but with different affinities. Experimental work performed by Dr. Marcella Ma.

5.6 p.A163T *FTO* variant is significantly associated with leanness in males from GWAS data

Analysis of data from the UK Biobank study (May 2016) revealed that the p.A163T variant was associated in males with a 0.8kg/m2 reduction in BMI (p=4x10-3). There was no detectable effect of this variant on puberty for AAM or voice breaking from this dataset.

5.7 *FTO* deficiency *in vivo* results in delayed vaginal opening in mice

In order to examine the influence of FTO activity on pubertal timing in an *in vivo* model, we examined timing of puberty in mice deficient for FTO in the heterozygous state ($Fto^{+/-}$). $Fto^{+/}$ mice had significantly delayed timing of vaginal opening (mean postnatal day +/- SEM: 27.20 +/- 0.44 in wild-type (n=31) vs 28.56 +/- 0.48 in $Fto^{+/}$ mice (n=45), p =0.042) (Fig. 5.7.1), an event which reflects the pubertal rise in estradiol (283).

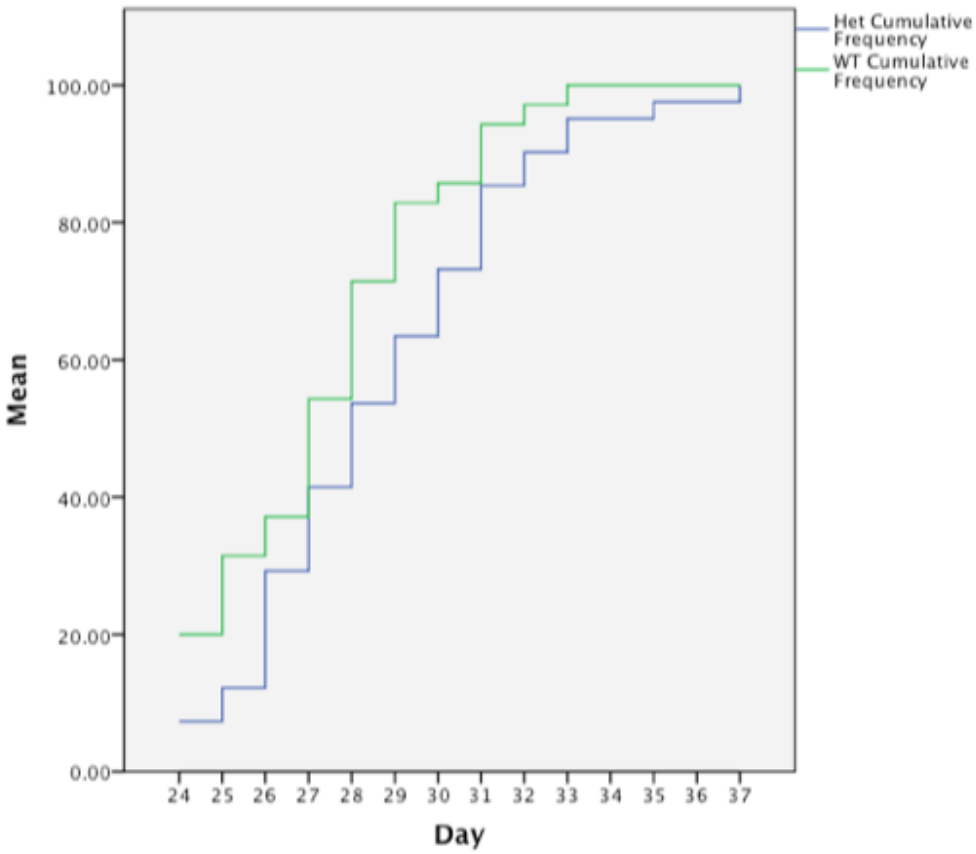


Figure 5.7.1 – Timing of vaginal opening in wild-type (WT) and *FTO*$^{+/-}$ heterozygous (Het) mice. Cumulative percentages of mice displaying vaginal opening by postnatal day are shown for WT and *FTO*$^{+/-}$ mice. WT mice n=31, *FTO*$^{+/-}$ n=45. p <0.05. Experimental work by Dr. Tony Coll and Dr. Debra Rimmington.

Mean body weight of the *Fto*$^{+/}$ group was lower than the wild type mice, but did not reach statistical significance (mean body weight (in g) +/- SEM: 11.64 +/- 0.21 in wild-type vs 11.45 +/- 0.14 in *Fto*$^{+/}$ mice, p=0.467) (Fig. 5.7.2).

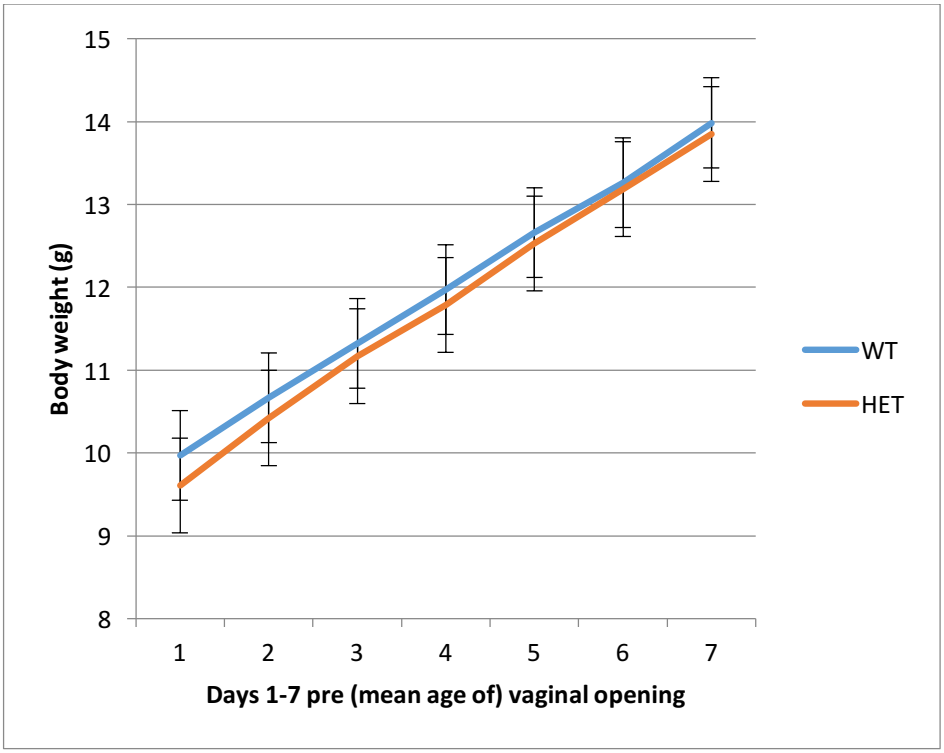


Figure 5.7.2 – Mean body weight (g) for wild type (WT) and *Fto*$^{+/-}$ (Het) mice in 7 days prior to vaginal opening. Mean body weight (in g) +/- SEM: 11.64 +/- 0.21 in wild-type vs 11.45 +/- 0.14 in *Fto*$^{+/}$ mice, p=0.467. Error bars show SEM for each group each day. Experimental work by Dr. Tony Coll and Dr. Debra Rimmington.

Using simple linear modelling (SPSS, version 24) to further analyse this data, *Fto* genotype of the pup (Het vs WT) explained approximately 3% of the total variation in puberty timing. Consideration of an additional factor, maternal genotype, improved the model by increasing the significance of the association between pup genotype and timing of puberty slightly (p=0.04), and accounted for 6% of the total variation in pubertal timing. In contrast, paternal genotype decreased the significance and total variation accounted for by the model.

5.8 Discussion

Genome wide association studies of age of menarche in the general population have attempted to unravel the complex conundrum of which genetic factors control the timing of puberty. Despite many loci being identified, clear evidence for the role of particular genes and pathways remains mostly lacking. Those genes lying within pathways of energy metabolism and growth appear promising, with the discovery of the role of *Lin28B* in C.elegans development (197) and the importance of leptin as a permissive signal in triggering the onset of puberty (284). Interestingly, self-limited or 'constitutional' DP is often associated with slow maturation throughout childhood, implicating growth and energy metabolic pathways in its pathogenesis. Previously, genes in such pathways identified through GWAS have not been screened in patients with DP.

The results from this study have implicated deleterious variants in *FTO* as a likely causal factor in the development of self-limited DP in three pedigrees from our large cohort of patients with familial DP. Rare variant burden testing identified potentially pathogenic variants in *FTO* to be significantly associated with self-limited DP. 2 rare variants in *FTO* were seen in 3 pedigrees with classic autosomal dominant segregation. *In silico* and *in vitro* testing of these 2 variants has not been able to define them as definitively pathogenic. However, the p.Leu44Val variant affects part of an alpha helix motif, which is highly conserved across placental mammals and may play a role in sexual function. Furthermore, this mutant protein displayed reduced kinetic activity *in vitro*.

In addition, recent data from the UK Biobank has revealed a significant association between the p.Ala163Thr variant identified in this study and a reduction in BMI in men. Notably, our patients identified with *FTO* variants showed reduced body mass, with individuals with the

p.Ala163Thr variant being particularly underweight. The *FTO* variants carried by our DP patients may result in reduced fat mass, which would in turn contribute to a delay in the timing of pubertal onset. It is possible that this delay in pubertal timing may be mediated directly through reduced leptin levels. Although we do not routinely measure leptin levels in our DP patients, leptin levels have been shown in published data to be significantly lower in pubertal-age patients with self-limited DP (284).

We have further explored the relationship between *FTO* heterozygosity and pubertal timing using an *in vivo* mouse model. $Fto^{+/-}$ mice had a significantly delayed onset of puberty as compared to WT mice. However, in the 7 days preceding puberty onset, body weight was not significantly different between the two pup genotype groups, although there was a trend towards the $Fto^{+/-}$ mice being lighter. Previous studies have demonstrated *Fto* genotype to have an influence on body weight. $Fto^{-/-}$ mice show a 30-40% reduction in body weight by 6 weeks of age (281). Transgenic mice with additional copies of *Fto* show a dose-dependent increase in body and fat mass [(285)]. Both $Fto^{-/-}$ and $Fto^{+/-}$ deficient mice show resistance to obesity, accumulate less white adipose tissue and have lower serum leptin levels than their *Fto*-replete litter-mates (281).

The relationship between FTO genotype, fat mass and leptin levels does remain somewhat unclear. Moreover, it is possible that *FTO* gene dosage may have an effect on energy homeostasis independent of effects on fat mass (285). FTO has been identified as an amino acid sensor and may play a role in developmental functions other than in body mass (286). FTO is expressed within the hypothalamus in several key sites critical for energy balance, including in the arcuate nucleus within proopiomelanocortin (POMC) neurons (287, 288). In one study, Fto levels in the arcuate nuclei of fasted mice fell by up to 60%, and this

reduction was not rescued by leptin administration (287). However, other studies have shown conflicting results in the effects on *Fto* mRNA levels of fasting, depending on whether whole hypothalamus or arcuate nucleus were studied and on the length of fast (288). However, *Fto*$^{-/-}$ mice have been shown to display blunted starvation-induced Npy mRNA induction (281). Thus, FTO may be important for signalling energy sufficiency and the 'healthy energy balance' required for pubertal onset. Indeed, our own *in silico* analysis has suggested that the p.Leu44Val mutation identified may represent a mammal-specific interaction site between FTO and another protein (or DNA), important for FTO function in terms of reproductive development. Moreover, the importance of maternal genotype on determining pubertal timing as demonstrated from our *Fto*$^{+/-}$ mice data is interesting. A reproductive phenotype present in *Fto* heterozygote mothers could expose pups to a suboptimal environment that could influence their puberty timing, independent of their own genotype.

Finally, our finding of maturational delay in growth in early childhood in the two probands with p.Leu44Val mutation is of interest. Constitutional delay in growth is seen in a subset of patients with DP, and our findings implicate mutations in energy homeostasis genes in the pathogenesis of patients with such a phenotype.

It is noteworthy that mutations in *FTO*, including the p.Ala163Thr variant identified in this study, have been identified in human subjects associated with both lean and obese phenotypes (251); however, the previous finding of obesity associated with the p.Ala163Thr variant may have been due to the study lacking power. We were not able in our study to identify the mechanism by which the p.Ala163Thr variant might lead to body mass reduction; although no reduction in demethylation activity was revealed using our assay it is

possible that this variant may produce a deleterious effect by another route, for example defects in post-translational modification or protein degradation.

In summary, we have identified two rare variants in *FTO* in our cohort of patients with self-limited DP, one of which, p.Ala163Thr, has for the first time been shown to be strongly associated with male leanness. Overall, our findings of a deleterious mutation in *FTO* in families with self-limited DP, and of delayed vaginal opening in FTO-deficient mice, provides evidence that perturbations in pathways of energy homeostasis and growth can produce a phenotype of DP. However, to date, our results suggest that overlap between those loci identified from GWAS of age at menarche, and genetic regulators in self-limited DP, is limited to a subset of patients. This hypothesis merits further exploration in our own cohort and in other populations.

Chapter 6

Results 4: Mutations in *IGSF1* may be associated with Self-Limited Delayed Puberty

6.1 Variants in *IGSF1* identified in DP families following targeted and whole exome sequencing

The *Immunoglobulin superfamily member 1* (*IGSF1*) gene encodes a plasma membrane glycoprotein mainly expressed in pituitary and testes. Loss-of-function mutations in the *IGSF1* gene have recently been described as causing a syndrome of X-linked central hypothyroidism, macro-orchidism, and DP (delayed rise of testosterone, but normal timing of testicular growth) (253). These mutations were found in a cohort of male patients with familial central hypothyroidism, but in some patients thyroid dysfunction was at a sub-clinical level. In most patients, prolactin deficiency and increased BMI were also observed. Heterozygous female carriers also exhibited delayed age at menarche (≥15 years) in 39%, central hypothyroidism in 33% of cases, prolactin deficiency in 11%, and increased body mass index in 61% (289). Whilst it is unknown whether *IGSF1* mutations can also cause isolated DP, recent GWAS have identified a locus close to *IGSF1* as significantly associated with both AAM and age at voice breaking in males (203).

Our cohort of families with self-limited DP was searched for families with an apparently maternally transmitted form of DP (pubertal delay in the index patient's siblings, mother, and/or siblings of mother) and initially Sanger sequencing analysis for mutations in *IGSF1* was performed on 30 male probands. Our WES results from the pipeline described above (Fig. 3.1.1) were additionally searched for variants in *IGSF1*.

Besides 13 polymorphisms in *IGSF1*, two variants were detected for which *in silico* analysis predicted possible pathogenicity (NM_001170961.1: c.3243G>C, (rs145402054) p.Met1081Ile, and NM_001170961.1: (rs146462069) c.1811A>C, p.Asn604Thr), and one variant with a low probability of pathogenicity (NM_001170961.1: c.2954T>C, (rs147496468) p.Val985Ala). All three variants had low minor allele frequencies in population databases. Details of the families in whom these variants were identified can be found in Joustra et al (290). In addition I identified one further family from our cohort (Family 4) with apparent maternally transmitted pattern of inheritance carrying this third variant p.Val985Ala in *IGSF1,* which segregated with the DP trait (Fig. 6.1.1).

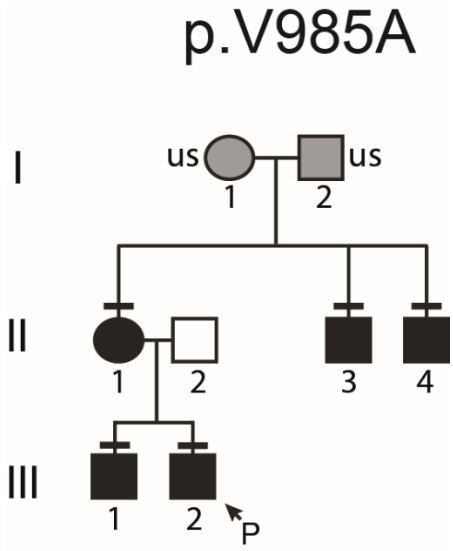


Figure 6.1 1- Pedigree of the family with *IGSF1* p.Val985Ala variant. Squares indicate male family members, circles female family members. Black symbols represent clinically affected, grey represent unknown phenotype, clear symbols represent unaffected individuals. The arrow with 'P' indicates the proband in each family. A horizontal black line above an individual's symbol indicates they are heterozygous for that mutation as confirmed by either whole exome sequencing or Fluidigm array, and verified by Sanger sequencing**.**

	Family 4 - c.2954T>C, p.Val985Ala (rs147496468), MAF/MAC 0.003/5				
Case	III.2 (proband)	II.2	II.3	II.4	III.1
Age, yr	33	72	70	56	35
Gender, M/F	M	F	M	M	M
***Pubertal development*[1]**					
Genital stage 2	16.02yrs	n.a.	-	-	-
Acceleration growth spurt	15.43yrs	-	-	>14.1yrs	14.02yrs
Peak height velocity	16.02yrs	?[4]	?[5]	?[6]	15.5yrs
Menarche, age (years)	n.a.	15	n.a.	n.a.	n.a.
FT4, pmol/L (10.0 – 23.0)	Within normal range	-	-	-	-
TSH, mU/L (0.8 – 6.0)	Within normal range	-	-	-	-
Prolactin, mU/L (85 – 319)	Within normal range	-	-	-	-
IGF-I, SDS[2] (-2.0 – 2.0)	Within normal range	-	-	-	-
Adult testis volume, mL[3] (6.7 – 20.8)	14.1	n.a.	-	-	-
BMI, kg/m^2 (18.0 – 25.0)	23.7	19.2	17.0	15.6	21.7

Table 6.1.1 - Clinical characteristics of our subjects with p.Val985Ala *IGSF1* variant.

Abbreviations: n.a., not applicable; MAF, minor allele frequency; MAC, minor allele count (according to 1000 Genome project, dbSNP build XML137); BMI, body mass index. [1]Z-score of age of occurrence. [2]Expressed in standard deviation scores, reference range from in-house normal values from the University Medical Center Utrecht, the Netherlands. [3]Volume of largest testis, measured with ultrasonography and calculated as π/6 x length x width x height. Reference range from Goede et al (291). [4]Grew 25cm after the age of 12.5 years. [5]Grew 38.5cm after the age of 12.5 years. [6]Grew 32cm after the age of 14.1 years

The proband in this family had a testicular volume of 7.3 mL at 16.5 years, in keeping with his DP, and his mother, brother and two of his mother's brothers all had DP with delayed linear growth. The p.Val985Ala variant was detected in this proband, his mother, brother and two of his mother's brothers. His father did not carry the variant and had normal timing of puberty. No other signs of *IGSF1* deficiency, such as such as central hypothyroidism, prolactin deficiency, or macro-orchidism, were present in the index case or other family members (Table 6.1.1).

6.2 *IGSF1* variants did not display pathological features on *in vitro* testing

Rare variants of unknown clinical significance (VUCSs) were tested *in vitro* for their effects on IGSF1 plasma membrane trafficking. Heterologous HEK293 cells were transfected with expression vectors specific for each *IGSF1* variant, followed by cell surface biotinylation and immunofluorescence to determine plasma membrane expression of the resulting IGSF1 proteins, as previously described (253). However, this p.Val985Ala variant exhibited normal expression at the plasma membrane when transfected into HEK293 cells (290), unlike other known pathogenic mutations in *IGSF1* (253).

6.3 Discussion

Whilst loss-of-function mutations in the *IGSF1* gene cause a syndrome of X-linked central hypothyroidism, macro-orchidism, and DP, pathogenic mutations in *IGSF1* have not so far been identified in our cohort of self-limited DP families. However, normal membrane trafficking of the IGSF1 protein cannot rule out possible functional effects of the observed *IGSF1* variants, since the function of various regions of the IGSF1 protein is unknown.

Moreover, it remains unclear by what mechanism loss-of-function mutations in *IGSF1* lead to the observed syndrome. In conclusion, whilst *IGSF1* mutations do not appear to be a common cause of isolated DP, it remains unknown whether *IGSF1* mutations can also cause self-limited DP, or contribute to its phenotype via a digenic or oligogenic inheritance pattern.

Chapter 7

Results 5: Mutations in further genes implicated in GnRH Neuronal Migration, Development and Function are found in patients with Self-Limited Delayed Puberty

7.1 – Genes identified from microarray data comparing expression in GnRH:GFP primary rat neurons during and after migration

In collaboration with Dr Anna Cariboni and her team at the University of Milan, microarray data from a comparative analysis of gene expression was examined. Firstly, we focused on gene expression in GnRH:GFP primary rat neurons at E14, E17 and E20 after cell sorting. The genes identified from this analysis that had significant up- (n=378) or down-regulation (n=299) during the time period of embryonic GnRH neuronal migration were shared with our group so that I could examine them for overlap with the 1765 genes with rare, predicted deleterious variants that segregated with trait within families that had been identified from the filtering of our DP whole exome sequencing data (Fig. 3.1.1). 48 genes were identified which were significantly up- or down-regulated (fold change > 2) between GnRH:GFP primary rat neurons at E14 and at E20, and that were also identified as potentially pathogenic in patients with self-limited DP (Fig. 7.1.1 and Table 7.1.1).

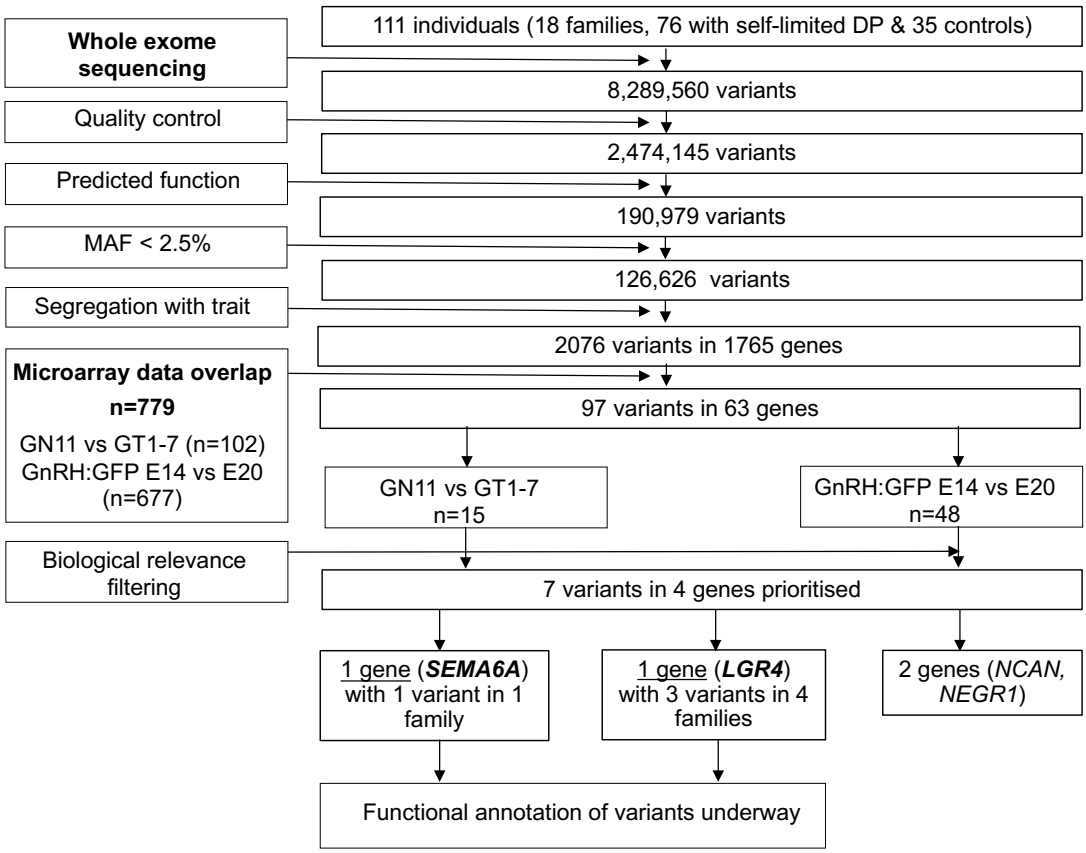


Figure. 7.1.1 – Filtering strategy for identification of candidate genes with overlap between comparative microarray analysis and whole exome sequencing in DP cohort. Biological relevance filtering see section 2.3.2.

Gene Symbol			
GNL2	KDSR	SCRN2	ERBB2
ASPM	CACNA1E	NCAN	CASP8AP2
GLI2	BPNT1	RPA1	CELF2
PBX2	KCNH7	APOA1BP	TPM1
SLCO3A1	CLASP2	CNTN5	CLDN9
ZC3H3	ETFDH	PRIM2	MADD
MED27	FYN	ZDHHC7	COL1A2
LGR4	PLXNA4	DPYSL3	GATA2
TIMELESS	DOLK	SULT1C3	JPH3
DNAAF3	LRRC55	UHRF1	GTF3C3
SKA3	FAT3	DSCAM	SNAP91
UBN1	DISP2	PTPRT	LRRN1

Table 7.1.1 – Genes identified from comparative analysis of GnRH:GFP primary rat neurons at E14 and at E20, and also identified as potentially pathogenic in patients with self-limited DP.

7.2 – Genes identified from microarray data comparing expression in GN11 and GT1-7 cells

Secondly, we examined microarray data from a comparative analysis of genes differentially expressed in GN11 (immature and migratory) (265) and GT1-7 (mature and non-migratory) (262) immortalised GnRH neurons. Only 15 genes reaching statistical significance for differential expression in these two cell lines were also identified in the list of 1765 genes with rare, predicted deleterious variants that segregated with trait within families that had been identified from the filtering of the whole exome sequencing data from our DP cohort (Fig. 7.1.1 and Table 7.2.1).

Gene Symbol	
HDHD3	NRXN1
FKBP10	OPRM1
EXO1	PSMD9
DAB1	RRP9
KCNJ3	SEMA6A
L1TD1	TFPI
MYO1D	PTGS2
NEGR1	

Table 7.2.1 – Genes identified from both the comparative analysis of GN11 vs GT1-7 cells, and identified as potentially pathogenic in patients with self-limited DP

Biological relevance filtering (for methods see section 2.3.2) identified one of these 15 genes, *SEMA6A*, as a highly interesting candidate. It was found to be upregulated in GN11, immature migratory GnRH-like neuronal cells, as compared to GT1-7 cells. *SEMA6A* belongs to the large family of semaphorins which are secreted or membrane-bound proteins that

mediate repulsive effects on axon pathfinding during nervous system development. *SEMA6A* is expressed in developing neural tissue and is known to be required for the proper development of the thalamocortical projection (292). *SEMA6A* belongs to a subfamily characterised by an extracellular semaphorin domain, a transmembrane domain and a long cytoplasmic tail. Members of this class are known to perform neuronal migration guidance functions (293). Interestingly, the length of the cytoplasmic tail, which includes a EVL-binding site in *SEMA6A*, suggests that it may also function as a receptor. Mutations in another member of the semaphorin family, *SEMA3A*, have been identified in both mouse models and human patients with KS (294, 295). Both the secreted class III semaphorins (e.g. SEMA3A) and the membrane-bound class VI semaphorins (e.g. SEMA6A) act as chemo-inhibitory molecules in neuronal development (296). *SEMA3A*, has been shown to signal through both Neuropilin 1 and 2 (NRP1 and 2) receptors (294), and studies have implicated *SEMA6A* as a further ligand for NRP2 (297).

The predicted deleterious variant identified in one pedigree from our DP cohort (NM_001300780.1 c.T1268C p.I423T) is rare, highly conserved, lies within the functional SEMA domain of the protein and shows good segregation with the DP trait (Table 7.2.2, Fig. 7.2.1 and Fig. 7.2.2).

Gene	AA Change	Phylo P (254)	SIFT (237)	PolyPhen2 (236)	LRT (255)	Mutation Taster (256)	FATHMM (257)	GERP ++ (258)	MAF (%) in ExAC database Finnish/All
SEMA6A	p.I423T	C	D	P	D	D	T	6.77	not seen

Table 7.2.2 – Details of prediction of variants according to web-based prediction software programs, conservation across species and minor allele frequency in study population and control cohorts. Minor allele frequency data for the control population was retrieved from the ExAC Browser (Exome Aggregation Consortium (ExAC), Cambridge, MA: http://exac.broadinstitute.org, accessed May 2015). C –conserved, D – deleterious, disease causing or damaging, P – possibly damaging, N – neutral

species	match	gene	aa	alignment
Human			423	LRTMVRYRLTKIAVDTAA GPYQNH
mutated	not conserved		423	TKTAVDTAA GPYQN
Ptroglodytes	all identical	ENSPTRG00000017162	423	TKIAVDTAA GPYQN
Mmulatta	not conserved	ENSMMUG00000012265	423	SKPCPDEFPDFISCSEN
Fcatus	no homologue			
Mmusculus	all identical	ENSMUSG00000019647	423	TKIAVDNAA GPYQN
Ggallus	all identical	ENSGALG00000002294	423	TKITVDSAA GPYKN
Trubripes	all identical	ENSTRUG00000016754	426	TRIVVDNEA GPYKN
Drerio	all identical	ENSDARG00000043028	422	TRIAVDNAA GPHNN
Dmelanogaster	no homologue			
Celegans	no homologue			
Xtropicalis	all identical	ENSXETG00000018720	423	TKIAVDTAA GPYQN

Figure 7.2.1 - Multiple sequence alignment (m.s.a.) for the *SEMA6A* p.I423T residue. M.s.a. was generated using MutationTaster (256). The p.I423T residue is highly conserved amongst different species, GERP score 6.77.

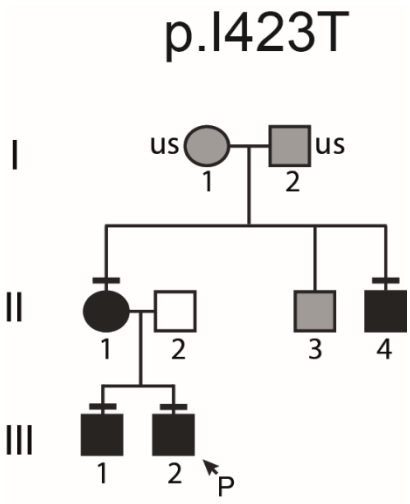


Figure 7.2.2 – Pedigree of the family with *SEMA6A* potentially pathogenic variant. Squares indicate male family members, circles female family members. Black symbols represent clinically affected, grey represent unknown phenotype, clear symbols represent unaffected individuals. The arrow with 'P' indicates the proband in each family. A horizontal black line above an individual's symbol indicates they are heterozygous for that mutation as confirmed by either whole exome sequencing or Fluidigm array, and verified by Sanger sequencing.

7.3 – Genes identified with overlap between comparative microarray analysis, whole exome sequencing in DP cohort and GWAS of age at menarche

The gene list generated from the overlap between the comparative microarray analysis (both GnRH:GFP primary rat neurons and GN11 vs GT1-7 cells) and the rare, predicted deleterious, well-segregating variants identified from the filtering of the whole exome sequencing data from our DP cohort was also matched with the 804 genes that had been identified as in linkage disequilibrium with loci identified with the largest and most recent GWAS of age at menarche (198). Only 4 genes remained as present in these overlapping lists (Fig. 7.1.2).

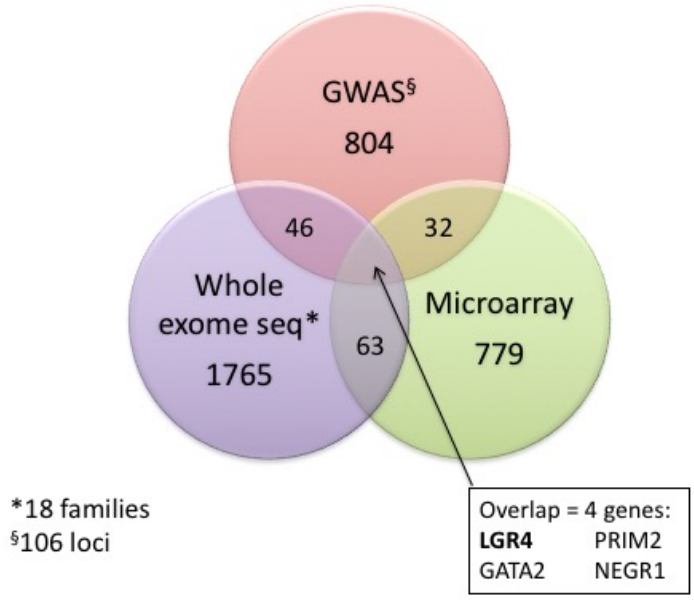


Figure 7.3.1 – Overlap between comparative microarray analysis, whole exome sequencing in DP cohort and GWAS of age at menarche.

The variants identified in *GATA2* and *PRIM2* in our DP cohort were found to be preferentially enriched in Finnish individuals from the ExAC database and on this basis they were de-prioritised. One potentially interesting variant was found in neuronal growth regulator *NEGR1* (p.Leu32Val). *NEGR1* is thought to be involved in cell adhesion and may act to promote axon sprouting in the mammalian brain. It has been implicated in the timing of age of menarche from GWAS of AAM (198).

However, most excitingly, 3 separate rare and highly conserved variants were identified in the G-protein coupled receptor *LGR4* in four families from our cohort (Fig. 7.3.1-7.3.5 and Table 7.3.1).

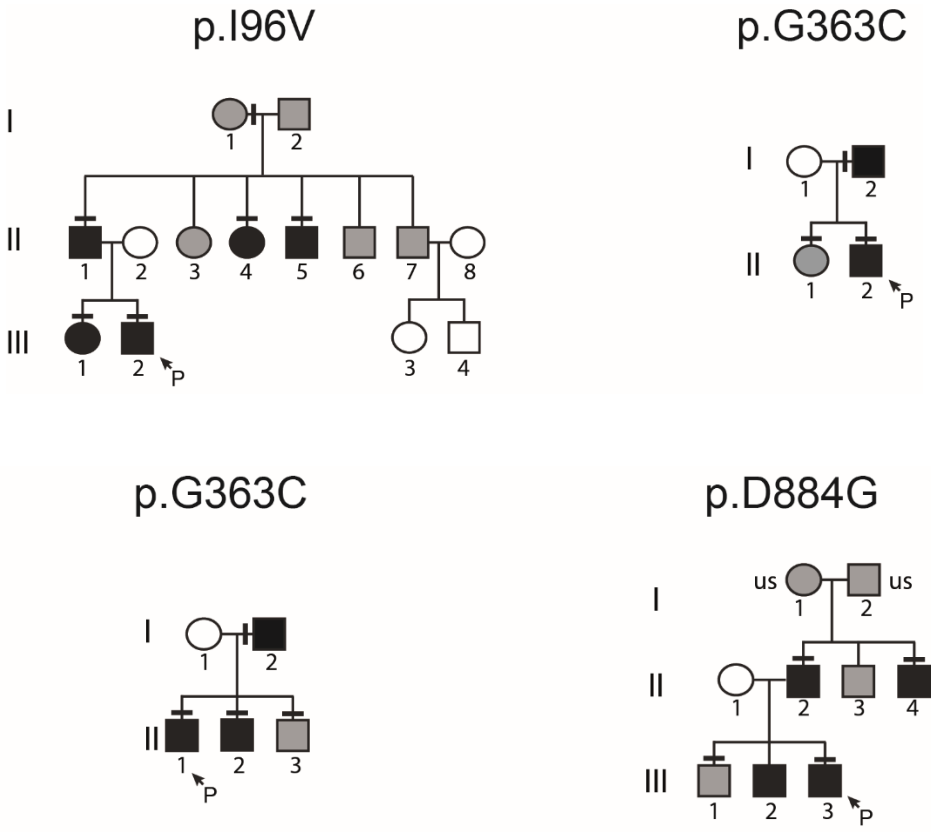


Figure 7.3.2 – Pedigrees of the families with *LGR4* potentially pathogenic variants

Squares indicate male family members, circles female family members. Black symbols represent clinically affected, grey represent unknown phenotype, clear symbols represent unaffected individuals. The arrow with 'P' indicates the proband in each family. A horizontal black line above an individual's symbol indicates they are heterozygous for that mutation as confirmed by either whole exome sequencing or Fluidigm array, and verified by Sanger sequencing.

LGR4 had already been identified as an interesting candidate gene along with *FTO* following rare variant burden testing of GWAS AAM genes with rare predicted pathogenic variants in DP patients (see Chapter 5 and Table 5.1.1).

Gene	AA Change	PhyloP (254)	SIFT (237)	PolyPhen 2 (236)	LRT (255)	Mutation Taster (256)	FATHMM (257)	GERP (258) ++	MAF (%) in ExAC database Finnish/All
LGR4	p.I96V	C	D	B	D	D	T	6.06	$0.04/8.8x10^{-5}$
LGR4	p.G363C	C	D	D	D	D	T	5.98	1.5/1.4
LGR4	p.D844G	C	D	P	D	D	T	5.39	0.18/1.2

Table 7.3.1 – Details of prediction of variants according to web-based prediction software programs, conservation across species and minor allele frequency in study population and control cohorts. Minor allele frequency data for the control population was retrieved from the ExAC Browser (Exome Aggregation Consortium (ExAC), Cambridge, MA: http://exac.broadinstitute.org, accessed May 2015). C –conserved, D – deleterious, disease causing or damaging, P – possibly damaging, N – neutral

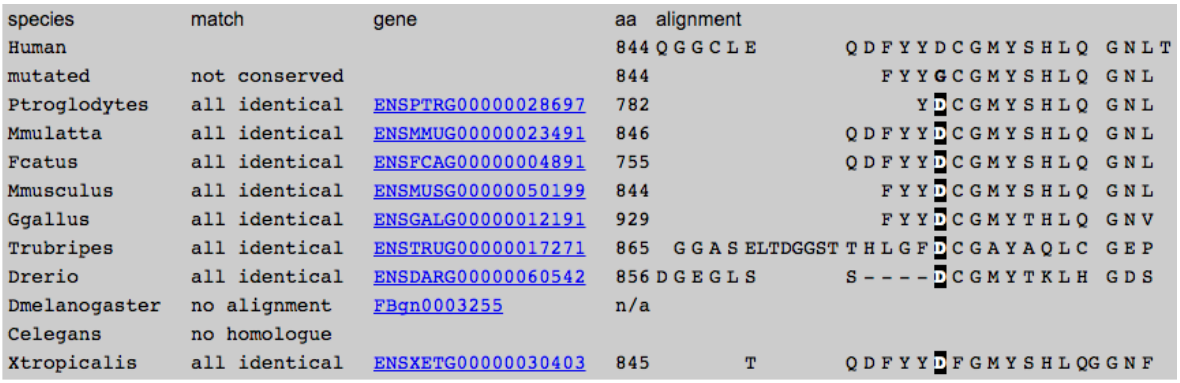
LGR4 codes for a receptor for R-spondins that potentiates the canonical Wnt signalling pathway and is involved in the formation of various organs, including eye, liver, reproductive tract and bone (298). One nonsense mutation in *LGR4* has also been shown to be associated with late onset of menarche (299). It is a large protein consisting of 18 extracellular LRRs together with a 7-transmembrane region (299). It is, therefore, a surface receptor highly homologous with *LGR5* and *LGR6*. *Lgr4* is highly expressed in the developing mouse embryo in both the vomeronasal organ and olfactory epithelium (300), pointing to a role in the development and/or migration of GnRH neurons.

species	match	gene	aa	alignment
Human			96	ELQLAGNDLSFIHPKALSGLKELK
mutated	all conserved		96	ELQLAGNDLSFVHPKALSGLKEL
Ptroglodytes	all identical	ENSPTRG00000028697	34	ELQLAGNDLSFIHPKALSGLKEL
Mmulatta	all identical	ENSMMUG00000023491	96	ELQLAGNDLSFIHPKALSGLKEL
Fcatus	no alignment	ENSFCAG00000004891	n/a	
Mmusculus	all identical	ENSMUSG00000050199	96	ELQLAGNDLSFIHPKALSGLKEL
Ggallus	all identical	ENSGALG00000012191	179	ELRLAGNDLSFIHPKALSGLKEL
Trubripes	not conserved	ENSTRUG00000017271	99	ELR-SGLHPRLFHTPLQPALSES
Drerio	all identical	ENSDARG00000060542	99	ELRLAGNDLAFIHPEALSGLHQL
Dmelanogaster	all conserved	FBgn0003255	218	DNSIINMDPNAFYGLAKL
Celegans	no homologue			
Xtropicalis	all identical	ENSXETG00000030403	98	ELRLAGNDLSIIHPMALSGLKEL

Figure 7.3.3 - Multiple sequence alignment (m.s.a.) for the *LGR4* p.I96V residue. M.s.a. was generated using MutationTaster (256). The p.I96V residue is highly conserved amongst different species, GERP score 6.06.

species	match	gene	aa	alignment
Human			363	YNNIRDLP SFNGCHALEEISLQRN
mutated	not conserved		363	FNCCHALEEISLQR
Ptroglodytes	all identical	ENSPTRG00000028697	301	GCHALEEISLQR
Mmulatta	all identical	ENSMMUG00000023491	363	FNGCHALEEISLQR
Fcatus	no alignment	ENSFCAG00000004891	n/a	
Mmusculus	all identical	ENSMUSG00000050199	363	FNGCRALEEISLQR
Ggallus	all identical	ENSGALG00000012191	446	YNNIKDLP SFTGCRSLEEISLQH
Trubripes	all identical	ENSTRUG00000017271	365	YNKITEIP TLQGCVRLQEINFQ
Drerio	all identical	ENSDARG00000060542	366	YNDIEDLP SFQGCVRLQDINLQ
Dmelanogaster	all identical	FBgn0003255	527	YNRIKALPQDAFQGIPKLQLLDLEGNEISYIHKEAFSGFTALEDLNLGN
Celegans	no homologue			
Xtropicalis	all identical	ENSXETG00000030403	363	V GFEGCSSLEEVYLQN

Figure 7.3.4 - Multiple sequence alignment (m.s.a.) for the *LGR4* p.G363C residue. M.s.a. was generated using MutationTaster (256). The p.G363C residue is highly conserved amongst different species, GERP score 5.98.



species	match	gene	aa	alignment
Human			844	QGGCLE QDFYYDCGMYSHLQ GNLT
mutated	not conserved		844	FYYGCGMYSHLQ GNL
Ptroglodytes	all identical	ENSPTRG00000028697	782	YDCGMYSHLQ GNL
Mmulatta	all identical	ENSMMUG00000023491	846	QDFYYDCGMYSHLQ GNL
Fcatus	all identical	ENSFCAG00000004891	755	QDFYYDCGMYSHLQ GNL
Mmusculus	all identical	ENSMUSG00000050199	844	FYYDCGMYSHLQ GNL
Ggallus	all identical	ENSGALG00000012191	929	FYYDCGMYTHLQ GNV
Trubripes	all identical	ENSTRUG00000017271	865	GGASELTDGGSTTHLGFDCGAYAQLC GEP
Drerio	all identical	ENSDARG00000060542	856	DGEGLS S----DCGMYTKLH GDS
Dmelanogaster	no alignment	FBgn0003255	n/a	
Celegans	no homologue			
Xtropicalis	all identical	ENSXETG00000030403	845	T QDFYYDFGMYSHLQGGNF

Figure 7.3.5 - Multiple sequence alignment (m.s.a.) for the *LGR4* p. D844G residue. M.s.a. was generated using MutationTaster (256). The p.D844G residue is highly conserved amongst different species, GERP score 5.39.

7.4 Discussion

Identification of biologically relevent candidate genes on which to focus functional annotation following next generation sequencing can be difficult, especially with more common traits such as DP. Comparison with other biological datasets, such as we have done here with microarray studies, can be a useful method to further filter the potentially pathogenic variants identified following WES.

Notably, misregulation of the migration of GnRH neurons during embryogenesis has been implicated in the pathogenesis of DP by our work on both *IGSF10* (269) and *HS6ST1*. Thus, expression data on genes which are up- or down-regulated during GnRH neuronal migration, or up- or down-regulated between immature migratory and mature GnRH-type neurons, are likely to be highly informative in the search for new candidates for the genetic basis of DP.

This comparative analysis has yielded several interesting new rare, potentially pathogenic variants in *SEMA6A*, *LGR4* and *NEGR1* in 7 families from our cohort so far; although more affected individuals with these variants or other potentially disease-causing variants in these genes may be identified by screening these genes more widely in our large cohort, and in reproduction studies in other cohorts. Functional annotation of all three of these genes to elucidate the mechanism by which they may contribute to the pathogenesis of DP is in progress or planned in our laboratory and the laboratory of Dr Cariboni's group in Milan.

Chapter 8: Conclusions and Future Directions

8.1 General Discussion

The genetic control of the timing of puberty remains an important but mostly unanswered question. Our lack of understanding of the factors that trigger pubertal onset are a barrier, to both the diagnosis and optimal management of patients with pubertal disorders. In addition, in order to comprehend the population-wide trend towards a more precocious age of pubertal onset in the developed world we must first gain a better understanding of puberty; both its physiological and genetic basis as a developmental process, and the mechanisms which elicit its onset. In view of the increasingly strong evidence for associations between early and late pubertal timing and poor health outcomes, these discoveries also clearly have a significant potential public health impact.

Attempts at unlocking this puzzle have ranged from population-based genome wide association studies, to whole exome and targeted sequencing in rare disorders of puberty. Together these studies suggest that pubertal timing in the general population may be controlled by hundreds of genetic regulators, whilst loss-of-function mutations in one gene can produce the phenotypic features of complete GnRH deficiency.

At the extreme end of the normal range of pubertal onset, self-limited DP is a common condition. Self-limited DP represents up to 65% of the children with DP presenting to Paediatric Endocrinologists. The inheritance of self-limited DP is known to be under strong genetic influence with clear autosomal dominant segregation of the trait within many families, and thus represents a useful basis for the investigation of puberty genetics.

However, as with other common complex diseases, identifying causal mutations in self-limited DP has proved difficult and for the majority of patients the genes responsible for self-limited DP have not been identified. To date only a small handful of genes have been implicated in this condition, mostly in a small number of relatives of patients with HH (77, 222, 223, 226, 227). Our large, accurately phenotyped cohort of patients with familial self-limited DP from Finland is a unique resource with a relatively homogeneous genetic composition. I have utilised this cohort to investigate the genetic variants segregating with trait within each pedigree. My hypothesis was that such families will be enriched for low frequency, high or moderate effect alleles that are amenable to discovery through exome sequencing.

Mechanistic candidates for the pathogenesis of DP, a type of functional hypogonadotropic hypogonadism, include genes that control the GnRH neuroendocrine network. Clearly, the phenotype of DP may be due to differences in factors affecting the 'pubertal brake' that is in place during early and mid-childhood, or, indeed, in changes to the timing of the release of this brake. We can postulate that these might include genetic differences such as upstream transcription factors and other genetic modifiers on the GnRH network and kisspeptin system, variations in permissive factors such as leptin and other indicators of healthy energy balance, and a host of environmental influences including body mass, exposure to endocrine disruptors and psycho-social problems.

However, as it remains unclear how normal pubertal onset is triggered in healthy individuals, it is difficult to narrow down the focus of our search only to events around the time of pubertal onset in early adolescence. Alterations in the underlying GnRH system resulting in its compromise, even during fetal life, thus leading to a system that is more, or

less, ‘sensitive’ to any affecters that act upon it, might also lead to a phenotype of DP. Such differences might affect GnRH neuronal development, activation or function, or downstream pathways including pituitary development and function. Whilst patients with HH caused by a neuronal migration defect may have a severally reduced number of GnRH neurons in place postnatally, a less drastic reduction in number might lead to a milder phenotype such as reversible HH or even DP. Moreover, given the complexity of the GnRH neuroendocrine network and the importance of neuronal synchrony, slow migration of GnRH neurons in embryonic life leading to a delay in arrival of a normal or near-normal number of neurons to the hypothalamus may also be damaging. Such late arrival may prevent the normal integration of these neurons into the neuroendocrine circuitry, resulting in a system that requires more stimulus to draw it out of its childhood dormancy; reflected in a clinical picture of DP. A similar hypothesis could be drawn for embryonic defects in GnRH action and its pituitary consequences. For example, whilst severe abnormalities such as a non-functional GnRH receptor result in HH, milder defects have been linked to both reversible HH and DP.

In view of the phenotypic overlap of self-limited DP with mild or reversing HH, I have used what is already know about the genetic basis of HH to inform the search for important genetic determinants of DP. Additionally, I have explored the wealth of data produced by GWAS of age of menarche to try to determine if there is any overlap between those variants controlling the timing of puberty in the general population and those leading to a phenotype of DP. Furthermore, via collaboration with another laboratory I have compared our whole exome data with results of microarray studies of embryonic GnRH neurons to try to pinpoint further important candidates for the genetic basis of DP.

My investigation to identify the causal genetic regulators in our large, well-phenotyped cohort of familial DP has uncovered several novel candidates. Through initial whole exome sequencing in eighteen highly informative families, and follow up targeted re-sequencing in an extended subgroup of the cohort, and using known and possible candidate genes from the published literature, I have discovered potential novel genetic regulators in self-limited DP. To date, whole exome and targeted re-sequencing in a total of 49 families has identified a likely causal mutation in 14 families from our cohort. These include: *HS6ST1*, mutations in which are already known to cause idiopathic hypothalamic hypogonadism; *FTO*, a candidate gene identified from GWAS and a known regulator of fat mass; and *IGSF10*, mutations in which have not previously been isolated in human disease.

IGSF10 emerged as a principal candidate gene following the analysis of my whole exome data via our in-house bioinformatics pipeline, including rare variant burden testing post targeted re-sequencing. I identified four rare heterozygous variants in *IGSF10* in 31 members of 10 unrelated families from our patient cohort. The identified variants are in evolutionarily conserved positions, predicted to have a deleterious effect on protein function and are significantly enriched in our DP cohort as compared to the general population ($p=0.020$). I carried out tissue expression studies, the results of which showed that *Igsf10* mRNA is strongly expressed in the nasal mesenchyme in mouse and human embryos, during the time-period when GnRH neurons migrate from their nasal origin towards the hypothalamus. *In vitro* analysis demonstrated that *Igsf10* knockdown caused reduced migration of immature GnRH neurons, and furthermore *Igsf10* knockdown by the use of morphilino oligonucleotides in a transgenic zebrafish model led to perturbed migration and extension of GnRH neurons.

These results pointed to a novel causal mechanism for self-limited DP, through the dysregulation of GnRH neuronal migration during embryonic development (Fig. 8.1). IGSF10 is likely to act as a signalling or axon-guidance protein, part of the complex system of secreted molecules and chemotactic gradients directing GnRH neuronal migration in the nasal region. Additionally, potentially pathogenic variants in *IGSF10* were then identified in 10% of patients with HH from a cohort collected collaboratively, including two patients with hypothalamic amenorrhoea. These results suggest that there may be a shared genetic basis between different types of functional hypogonadotropic hypogonadism, including both self-limited DP and hypothalamic amenorrhoea (or its male-equivalent).

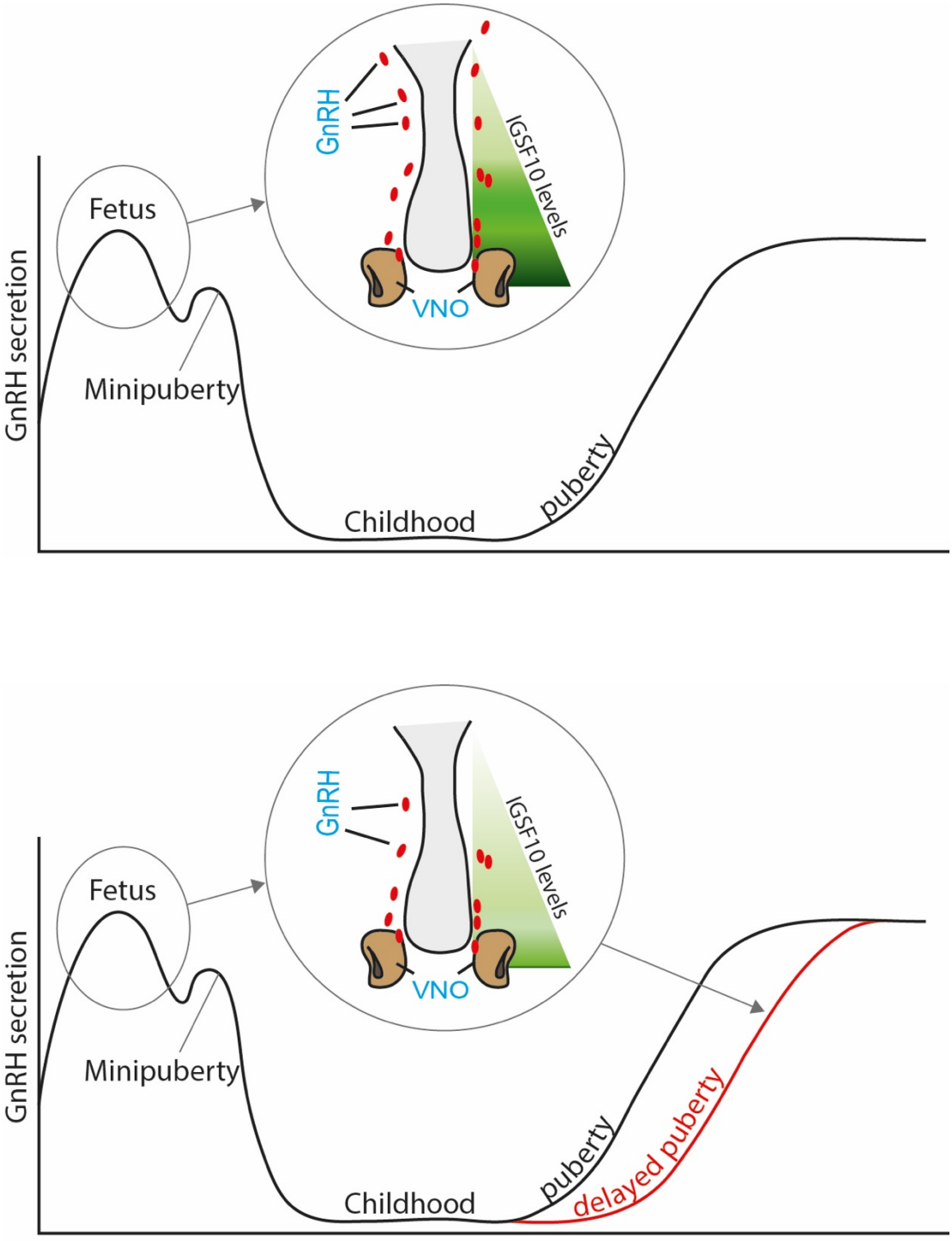


Figure 8.1 – Schematic for the role of IGSF10 in the late timing of pubertal onset.

In the lower panel, as compared to the top panel, a reduced gradient of IGSF10 signal leads to dysregulation of GnRH neuronal migration during fetal life, impacting on the timing of puberty in late childhood.

On searching our filtered whole exome data specifically for genes known to be related to the phenotype of HH, a novel variant in *HS6ST1* was found in six affected members of one family from the cohort. Reduced enzymatic activity of the mutant protein resulting from this missense variant was demonstrated by the use of a sulphotransferase assay. *Hs6st1* was shown to be expressed both at sites of GnRH neuronal development in the vomeronasal organ of the embryonic nasal placode, and within the peri-pubertal hypothalamus in mice.

This data suggests that a *HS6ST1* mutation leading to a milder reduction in sulphotransferase activity than that seen with HH-causing mutations may produce a DP phenotype. Moreover, the expression pattern of *Hs6St1* lends further weight to the hypothesis formed from the discovery of *IGSF10* as a DP candidate gene, that perturbations in the development or migration of GnRH neurons, perhaps relativity subtle in nature, may lead to 'simple' DP, whilst more deleterious alterations will result in severe phenotypes such as HH. Interestingly, no other pathogenic mutations in genes known to cause HH have been identified in our cohort to date, suggesting that the genetic background of DP may be significantly different to that of HH, or shared by as yet undiscovered genes.

My analysis for overlap between GWAS of AAM and our cohort data identified two missense variants in *FTO* in 3 families. The demethylase activity of the mutant proteins resulting from these variants have been quantified using an established assay. Mice studies showed a significantly delayed timing of vaginal opening, a marker of pubertal onset, in *FTO* heterozygote mice as compared to wild-type, providing an *in vivo* model of the effects of reduced *FTO* gene dosage on pubertal timing.

Whether the effects of *FTO* on pubertal timing are dependent or independent of body mass remains unclear. One of the variants identified in our cohort is strongly associated with male

leanness. Our human subjects display both significant thinness as compared to normalized data and delayed growth, some from early childhood. The heterozygote mice studies also showed a trend towards reduced body mass as compared to their wild-type controls, although this did not meet significance. However, our findings suggest that mutations in genes such as *FTO*, which affect or are signals of energy homeostasis and growth, can contribute to the phenotype of DP. To date, the evidence for overlap between those loci identified from GWAS of age at menarche representing potential genetic regulators of timing of puberty in the general population, and genetic regulators in self-limited DP, remains limited.

Three further candidate genes, *SEMA6A*, *LGR4* and *NEGR1*, were idenitifed with potentially pathogenic variants in 7 families from our cohort via a comparitive study of our filtered whole exome sequencing data and microarray analysis of genes which are up- or down-regulated during GnRH neuronal migration. All 3 are promising candidates for potential regulators of GnRH development or migration and will be taken forward for further study. Finally, I have analysed our data for potentially pathogenic variants in another member of the IGSF superfamily, *IGSF1*. Although one variant was identified from our whole exome sequencing data, the *in vitro* analysis did not suggest that this variant was likely to be a causal mutation in our patients.

It is interesting to see that several of our large pedigrees with whole exome data available do not have an obvious mutation so far identified. Indeed, many families in our cohort remain without apparent genetic mutations that would explain their phenotype, despite clear autosomal dominance inheritance patterns in those pedigrees. I have encountered several difficulties in the identification of clear casual mutations in affected individuals

during this work. One of the principal difficulties has been in the filtering of the huge amount of genetic data returned from our whole exome sequencing in 18 families. Whole exome sequencing has been of great value in identifying novel causal mutations in patients with highly rare diseases with autosomal recessive inheritance. However, as self-limited DP is present in 2-2.5% of the normal population, individuals with DP are predicted to make up a similar proportion of any sequencing database, thus preventing filtering of non-novel variants during analysis. This factor also increases the requirement for determining statistical significance by methods such as whole gene rare variant burden testing, to identify genes that have an enrichment of potentially pathogenic variants in patients in comparison to control subjects from the public databases. Additionally, the autosomal dominant or bilineal inheritance pattern in the majority of families has meant that variants that are inherited in a heterozygous manner cannot be discounted from the analysis; indeed, they are expected. These factors are in addition to the standard issues of whole exome sequencing analysis: of identifying true variants rather than sequencing artefacts via quality control, prediction of which variants are likely to affect protein function or be truly 'pathogenic', verification of variants through Sanger sequencing and reproduction of results in secondary datasets.

The difficulties in identifying key casual variants in the majority of pedigrees screened may suggest that in these families there is more than one genetic factor at play; that perhaps the 'low hanging fruit' has been found and the next steps will be more complex. Indeed, the question of how to identify digenic and oligogenic inheritance in familial disease is being asked by many investigators in many different disciplines at the present time. Furthermore, in a condition such as DP there may be genetic factors causing late puberty, as well as

environmental factors (e.g. increased BMI) causing early puberty in direct competition at play in the same individual. These difficult issues highlight the importance of large cohorts, with accurate phenotyping, as the basis for identification of causal mutations in a condition such as DP.

A further consideration is the decision in my project to use whole exome sequencing as a gene identification tool, rather than whole genome sequencing. It is, of course, highly likely that some of the genetic factors regulating the timing of pubertal onset will be found in the non-coding regions of the genome, and whole exome sequencing will fail to identify a large proportion of these variants. However, although the cost of genome sequencing is continuing to fall, its expense still precludes the sequencing of large numbers of individuals. Furthermore, the correct analysis and filtering of whole genome sequencing data is still unclear; vast amounts of data is generated and most of the variants cannot be annotated by current prediction software tools. Thus, until the understanding of whole genome sequencing has progressed and the costs have further reduced this remains a future plan for investigation of this cohort.

Finally, whilst understanding the genetic and biological control of the onset of puberty is in itself a fascinating pursuit, at the centre of this work remain the patients who are affected by pubertal abnormalities and the drive to inform their management through cutting-edge research. Currently, the differential diagnosis between CHH and DP in early adolescence remains challenging, as both diseases present with isolated delay in puberty. Moreover, there are no specific biochemical markers to accurately differentiate between these two disorders. Defining the genetic architecture of DP, as other groups have done and are

continuing to do for HH, is the first step towards using genetic testing to improving the accuracy and speed of diagnosis and management of young people with pubertal delay.

8.2 Future and Ongoing Work

In order to increase the power of our study with respect to whole gene rare variant burden testing, whole exome sequencing of a further 49 probands from our cohort has recently been completed, but is yet to be fully analysed. Together with Dr Claudia Cabrera (QMUL), I am modifying our in-house bioinformatics pipeline to allow for prioritisation of variants by rare variant burden testing immediately post quality controls filters, to ensure all variants prioritised meet stringent statistical significance thresholds. These additional probands will also allow a more comprehensive comparison of the overlap between genetic regulators of self-limited DP and both the timing of puberty in the general population (from GWAS of AAM) and genes causing HH; as this comparison will be based on a total of 67 families rather than the previous 18.

In addition, we have commenced 2 new collaborative projects. In collaboration with Prof. Yee Ming-Chan's group in Boston, USA, we are carrying out a combined analysis of the whole exome data from our two cohorts of patients with self-limited DP. The exomes will be called and analysed jointly, and comprise a total of 130 probands to further increase the power of our study. Moreover, genes with a significant burden of potentially pathogenic variants found both in our Finnish population and in patients collected in Boston represent more robust candidates to take forward for further study as they have been identified in two distinct populations.

Secondly, we are collaborating with Prof. Nelly Pitteloud's group from Lausanne. This group has extensive experience with identification of novel causal genes in their large cohort of patients with HH. Via a joint analysis of the Pitteloud group's whole exome data from HH patients and our self-limited DP data, we can compare the genetic architecture of HH and DP, and investigate whether genetic testing could assist in the differential diagnosis of young people presenting with functional hypogonadotropic hypogonadism.

Additionally, during this project I have sought and had granted ethical approval for the collection of a further cohort of patients with abnormal pubertal timing within the UK. To date I have collated DNA samples from a group of 20 probands with severely delayed onset of puberty, 5 probands with absent pubertal onset, and a small group of patients with precocious puberty. Where possible, DNA samples have also been collected from the affected and unaffected relatives of these probands. The aim of this cohort is to allow further genetic discovery through whole and targeted exome sequencing, to allow verification of findings identified through our Finnish DP cohort, and to assist the management of these UK patients. So far, only six probands have been analysed via WES. Of these six, two patients with severe and difficult to manage DP, but without clear red flag markers of HH, have been diagnosed after WES. One adolescent male patient has been identified as having a *GNRHR* defect known to result in partial HH, and an adolescent female patient has been found to be homozygous for a known mutation in *TAC3*, leading to an HH phenotype.

Through collaboration with Prof Ulrich Boehm's group we have plans to produce a conditional knockout murine model of *IGSF10.* No such model currently exists and no embryonic stem cell lines are available from either the Wellcome Trust Sanger Institute's

Mouse Genetics project or the European Conditional Mouse Mutagenesis Program. A mouse model would be invaluable for studies to examine timing of puberty in both total and heterozygous knockout mice, as well as providing tissue for *in vitro* work to characterise GnRH neuronal migration development and function in the absence of or with significantly reduced levels of *IGSF10.*

With the aim of further understanding the mechanism of action of the IGSF10 protein, we plan to study potential interactors and carry out experiments to attempt to identify the receptor or receptors through which IGSF10 mediates its signalling to GnRH neurons. Several options are being explored, including ligand-based receptor capture, affinity labelling or affinity chromatography, *in vitro* binding assays to putative transmembrane proteins and co-immunoprecipitation experiments.

Functional annotation of 2 variants in *EAP1*, and three *LGR4* variants, is currently underway by another member of our laboratory group, Alessandra Mancini, as part of her PhD project. This work has already been presented at an international meeting. The functional *in vitro* and *in vivo* work on the variants idenitifed in *SEMA6A* is being carried out in a collaboration with our group by Dr Anna Cariboni's laboratory in Milan. Functional work on the *NEGR1* variant is also planned.

8.3 Conclusion

My findings to date serve to highlight the fascinating heterogeneity of genetic mis-regulation resulting in disorders of pubertal timing such as self-limited DP. This work also points to the possibility of multiple pathophysiological pathways converging in a phenotype of delayed puberty. Already, the genetic factors implicated from this study influence the development of GnRH neuronal pathways (both migration and maturation) and body mass and metabolism. The continued search for answers in these families requires consideration of digenic and oligogenic inheritance, and exploration of other likely gene networks include upstream factors that contribute to the transcriptional control of the GnRH neuronal network, biological regulation of kisspeptin, and as yet undiscovered pathways. Clearly there is much ongoing work to be done and many more genetic players to be uncovered in the genetic control of self-limited DP.

References

1. Cariboni A, Maggi R, Parnavelas JG. From nose to fertility: the long migratory journey of gonadotropin-releasing hormone neurons. Trends in neurosciences. 2007;30(12):638-44.
2. Beate K, Joseph N, Nicolas de R, Wolfram K. Genetics of isolated hypogonadotropic hypogonadism: role of GnRH receptor and other genes. International journal of endocrinology. 2012;2012:147893.
3. Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. Archives of disease in childhood. 1970;45(239):13-23.
4. Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. Archives of disease in childhood. 1969;44(235):291-303.
5. Palmert MR, Dunkel L. Clinical practice. Delayed puberty. N Engl J Med. 2012;366(5):443-53.
6. Parent AS, Franssen D, Fudvoye J, Gerard A, Bourguignon JP. Developmental variations in environmental influences including endocrine disruptors on pubertal timing and neuroendocrine control: Revision of human observations and mechanistic insight from rodents. Frontiers in neuroendocrinology. 2015;38:12-36.
7. van Buuren S. Growth charts of human development. Stat Methods Med Res. 2013;23(4):346-68.
8. Bourguignon JP. Growth and timing of puberty: reciprocal effects. Hormone research. 1991;36(3-4):131-5.
9. Tanner JM. Growth and endocrinology of the adolescent. In: Gardner LI, editor. Endocrine and Genetic Diseases of Childhood and Adolescents. second ed. Philadelphia: WB Saunders; 1975.
10. Lawaetz JG, Hagen CP, Mieritz MG, Blomberg Jensen M, Petersen JH, Juul A. Evaluation of 451 Danish boys with delayed puberty: diagnostic use of a new puberty nomogram and effects of oral testosterone therapy. The Journal of clinical endocrinology and metabolism. 2015;100(4):1376-85.
11. Wagner IV, Sabin MA, Pfaffle RW, Hiemisch A, Sergeyev E, Korner A, et al. Effects of obesity on human sexual development. Nature reviews Endocrinology. 2012;8(4):246-54.

12. Gasser T, Sheehy A, Molinari L, Largo RH. Growth of early and late maturers. Ann Hum Biol. 2001;28(3):328-36.
13. Biro FM, McMahon RP, Striegel-Moore R, Crawford PB, Obarzanek E, Morrison JA, et al. Impact of timing of pubertal maturation on growth in black and white female adolescents: The National Heart, Lung, and Blood Institute Growth and Health Study. The Journal of pediatrics. 2001;138(5):636-43.
14. Gamba M, Pralong FP. Control of GnRH neuronal activity by metabolic factors: the role of leptin and insulin. Molecular and cellular endocrinology. 2006;254-255:133-9.
15. Juul A, Teilmann G, Scheike T, Hertel NT, Holm K, Laursen EM, et al. Pubertal development in Danish children: comparison of recent European and US data. International journal of andrology. 2006;29(1):247-55; discussion 86-90.
16. Teilmann G, Pedersen CB, Skakkebaek NE, Jensen TK. Increased risk of precocious puberty in internationally adopted children in Denmark. Pediatrics. 2006;118(2):e391-9.
17. Mul D, Fredriks AM, van Buuren S, Oostdijk W, Verloove-Vanhorick SP, Wit JM. Pubertal development in The Netherlands 1965-1997. Pediatric research. 2001;50(4):479-86.
18. Parent AS, Teilmann G, Juul A, Skakkebaek NE, Toppari J, Bourguignon JP. The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. Endocrine reviews. 2003;24(5):668-93.
19. de Muinich Keizer SM, Mul D. Trends in pubertal development in Europe. Human reproduction update. 2001;7(3):287-91.
20. Sorensen K, Aksglaede L, Petersen JH, Juul A. Recent changes in pubertal timing in healthy Danish boys: associations with body mass index. The Journal of clinical endocrinology and metabolism. 2010;95(1):263-70.
21. Herman-Giddens ME, Steffes J, Harris D, Slora E, Hussey M, Dowshen SA, et al. Secondary sexual characteristics in boys: data from the Pediatric Research in Office Settings Network. Pediatrics. 2012;130(5):e1058-68.
22. Euling SY, Herman-Giddens ME, Lee PA, Selevan SG, Juul A, Sorensen TI, et al. Examination of US puberty-timing data from 1940 to 1994 for secular trends: panel findings. Pediatrics. 2008;121 Suppl 3:S172-91.

23. Ong KK, Ahmed ML, Dunger DB. Lessons from large population studies on timing and tempo of puberty (secular trends and relation to body size): the European trend. Molecular and cellular endocrinology. 2006;254-255:8-12.
24. Biro FM, Khoury P, Morrison JA. Influence of obesity on timing of puberty. International journal of andrology. 2006;29(1):272-7; discussion 86-90.
25. He Q, Karlberg J. Bmi in childhood and its association with height gain, timing of puberty, and final height. Pediatric research. 2001;49(2):244-51.
26. Frisch RE, McArthur JW. Menstrual cycles: fatness as a determinant of minimum weight for height necessary for their maintenance or onset. Science. 1974;185(4155):949-51.
27. De Leonibus C, Marcovecchio ML, Chiarelli F. Update on statural growth and pubertal development in obese children. Pediatr Rep. 2012;4(4):e35.
28. Lee JM, Wasserman R, Kaciroti N, Gebremariam A, Steffes J, Dowshen S, et al. Timing of Puberty in Overweight Versus Obese Boys. Pediatrics. 2016;137(2):1-10.
29. Elias CF. Leptin action in pubertal development: recent advances and unanswered questions. Trends in endocrinology and metabolism: TEM. 2012;23(1):9-15.
30. Ahmed ML, Ong KK, Morrell DJ, Cox L, Drayer N, Perry L, et al. Longitudinal study of leptin concentrations during puberty: sex differences and relationship to changes in body composition. The Journal of clinical endocrinology and metabolism. 1999;84(3):899-905.
31. Barash IA, Cheung CC, Weigle DS, Ren H, Kabigting EB, Kuijper JL, et al. Leptin is a metabolic signal to the reproductive system. Endocrinology. 1996;137(7):3144-7.
32. Zuure WA, Quennell JH, Anderson GM. Leptin Responsive and GABAergic Projections to the Rostral Preoptic Area in Mice. Journal of neuroendocrinology. 2015.
33. Bellefontaine N, Chachlaki K, Parkash J, Vanacker C, Colledge W, d'Anglemont de Tassigny X, et al. Leptin-dependent neuronal NO signaling in the preoptic hypothalamus facilitates reproduction. The Journal of clinical investigation. 2014;124(6):2550-9.
34. Qiu X, Dao H, Wang M, Heston A, Garcia KM, Sangal A, et al. Insulin and Leptin Signaling Interact in the Mouse Kiss1 Neuron during the Peripubertal Period. PloS one. 2015;10(5):e0121974.
35. Pralong FP. Insulin and NPY pathways and the control of GnRH function and puberty onset. Molecular and cellular endocrinology. 2010;324(1-2):82-6.

36. Pomerants T, Tillmann V, Jurimae J, Jurimae T. Relationship between ghrelin and anthropometrical, body composition parameters and testosterone levels in boys at different stages of puberty. J Endocrinol Invest. 2006;29(11):962-7.
37. Persson I, Ahlsson F, Ewald U, Tuvemo T, Qingyuan M, von Rosen D, et al. Influence of perinatal factors on the onset of puberty in boys and girls: implications for interpretation of link with risk of long term diseases. American journal of epidemiology. 1999;150(7):747-55.
38. Adair LS. Size at birth predicts age at menarche. Pediatrics. 2001;107(4):E59.
39. Wehkalampi K, Hovi P, Dunkel L, Strang-Karlsson S, Jarvenpaa AL, Eriksson JG, et al. Advanced pubertal growth spurt in subjects born preterm: the Helsinki study of very low birth weight adults. The Journal of clinical endocrinology and metabolism. 2011;96(2):525-33.
40. Dunger DB, Ahmed ML, Ong KK. Early and late weight gain and the timing of puberty. Molecular and cellular endocrinology. 2006;254-255:140-5.
41. Ibanez L, Valls C, Ong K, Dunger DB, de Zegher F. Metformin therapy during puberty delays menarche, prolongs pubertal growth, and augments adult height: a randomized study in low-birth-weight girls with early-normal onset of puberty. The Journal of clinical endocrinology and metabolism. 2006;91(6):2068-73.
42. Aksglaede L, Juul A, Olsen LW, Sorensen TI. Age at puberty and the emerging obesity epidemic. PloS one. 2009;4(12):e8450.
43. Aksglaede L, Sorensen K, Petersen JH, Skakkebaek NE, Juul A. Recent decline in age at breast development: the Copenhagen Puberty Study. Pediatrics. 2009;123(5):e932-9.
44. Mouritsen A, Aksglaede L, Sorensen K, Mogensen SS, Leffers H, Main KM, et al. Hypothesis: exposure to endocrine-disrupting chemicals may interfere with timing of puberty. Int J Androl. 2010;33(2):346-59.
45. Caserta D, Maranghi L, Mantovani A, Marci R, Maranghi F, Moscarini M. Impact of endocrine disruptor chemicals in gynaecology. Human reproduction update. 2008;14(1):59-72.
46. Krstevska-Konstantinova M, Charlier C, Craen M, Du Caju M, Heinrichs C, de Beaufort C, et al. Sexual precocity after immigration from developing countries to Belgium: evidence of previous exposure to organochlorine pesticides. Hum Reprod. 2001;16(5):1020-6.

47. van den Driesche S, Macdonald J, Anderson RA, Johnston ZC, Chetty T, Smith LB, et al. Prolonged exposure to acetaminophen reduces testosterone production by the human fetal testis in a xenograft model. Sci Transl Med. 2015;7(288):288ra80.
48. Main KM, Skakkebaek NE, Virtanen HE, Toppari J. Genital anomalies in boys and the environment. Best Pract Res Clin Endocrinol Metab. 2010;24(2):279-89.
49. Swan SH, Main KM, Liu F, Stewart SL, Kruse RL, Calafat AM, et al. Decrease in anogenital distance among male infants with prenatal phthalate exposure. Environ Health Perspect. 2005;113(8):1056-61.
50. Dean A, van den Driesche S, Wang Y, McKinnell C, Macpherson S, Eddie SL, et al. Analgesic exposure in pregnant rats affects fetal germ cell development with inter-generational reproductive consequences. Sci Rep. 2016;6:19789.
51. Wehkalampi K, Widen E, Laine T, Palotie A, Dunkel L. Patterns of inheritance of constitutional delay of growth and puberty in families of adolescent girls and boys referred to specialist pediatric care. The Journal of clinical endocrinology and metabolism. 2008;93(3):723-8.
52. Gajdos ZK, Hirschhorn JN, Palmert MR. What controls the timing of puberty? An update on progress from genetic investigation. Current opinion in endocrinology, diabetes, and obesity. 2009;16(1):16-24.
53. Morris DH, Jones ME, Schoemaker MJ, Ashworth A, Swerdlow AJ. Familial concordance for age at menarche: analyses from the Breakthrough Generations Study. Paediatric and perinatal epidemiology. 2011;25(3):306-11.
54. Tena-Sempere M. Kisspeptin/GPR54 system as potential target for endocrine disruption of reproductive development and function. International journal of andrology. 2010;33(2):360-8.
55. Wray S. From nose to brain: development of gonadotrophin-releasing hormone-1 neurones. Journal of neuroendocrinology. 2010;22(7):743-53.
56. Akutsu S, Takada M, Ohki-Hamazaki H, Murakami S, Arai Y. Origin of luteinizing hormone-releasing hormone (LHRH) neurons in the chick embryo: effect of the olfactory placode ablation. Neuroscience letters. 1992;142(2):241-4.
57. Forni PE, Taylor-Burds C, Melvin VS, Williams T, Wray S. Neural crest and ectodermal cells intermix in the nasal placode to give rise to GnRH-1 neurons, sensory neurons, and

olfactory ensheathing cells. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2011;31(18):6915-27.

58. Herbison AE, Porteous R, Pape JR, Mora JM, Hurst PR. Gonadotropin-releasing hormone neuron requirements for puberty, ovulation, and fertility. Endocrinology. 2008;149(2):597-604.

59. Cariboni A, Rakic S, Liapi A, Maggi R, Goffinet A, Parnavelas JG. Reelin provides an inhibitory signal in the migration of gonadotropin-releasing hormone neurons. Development. 2005;132(21):4709-18.

60. Tobet SA, Schwarting GA. Minireview: recent progress in gonadotropin-releasing hormone neuronal migration. Endocrinology. 2006;147(3):1159-65.

61. Gao C, Noden DM, Norgren RB, Jr. LHRH neuronal migration: heterotypic transplantation analysis of guidance cues. J Neurobiol. 2000;42(1):95-103.

62. Cariboni A, Hickok J, Rakic S, Andrews W, Maggi R, Tischkau S, et al. Neuropilins and their ligands are important in the migration of gonadotropin-releasing hormone neurons. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2007;27(9):2387-95.

63. Schwanzel-Fukuda M, Abraham S, Crossin KL, Edelman GM, Pfaff DW. Immunocytochemical demonstration of neural cell adhesion molecule (NCAM) along the migration route of luteinizing hormone-releasing hormone (LHRH) neurons in mice. The Journal of comparative neurology. 1992;321(1):1-18.

64. Bulow HE, Berry KL, Topper LH, Peles E, Hobert O. Heparan sulfate proteoglycan-dependent induction of axon branching and axon misrouting by the Kallmann syndrome gene kal-1. Proceedings of the National Academy of Sciences of the United States of America. 2002;99(9):6346-51.

65. Magni P, Dozio E, Ruscica M, Watanobe H, Cariboni A, Zaninetti R, et al. Leukemia inhibitory factor induces the chemomigration of immortalized gonadotropin-releasing hormone neurons through the independent activation of the Janus kinase/signal transducer and activator of transcription 3, mitogen-activated protein kinase/extracellularly regulated kinase 1/2, and phosphatidylinositol 3-kinase/Akt signaling pathways. Molecular endocrinology. 2007;21(5):1163-74.

66. Giacobini P, Messina A, Wray S, Giampietro C, Crepaldi T, Carmeliet P, et al. Hepatocyte growth factor acts as a motogen and guidance signal for gonadotropin

hormone-releasing hormone-1 neuronal migration. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2007;27(2):431-45.
67. Corradi A, Croci L, Broccoli V, Zecchini S, Previtali S, Wurst W, et al. Hypogonadotropic hypogonadism and peripheral neuropathy in Ebf2-null mice. Development. 2003;130(2):401-10.
68. Schwarting GA, Henion TR, Nugent JD, Caplan B, Tobet S. Stromal cell-derived factor-1 (chemokine C-X-C motif ligand 12) and chemokine C-X-C motif receptor 4 are required for migration of gonadotropin-releasing hormone neurons to the forebrain. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2006;26(25):6834-40.
69. Soussi-Yanicostas N, de Castro F, Julliard AK, Perfettini I, Chedotal A, Petit C. Anosmin-1, defective in the X-linked form of Kallmann syndrome, promotes axonal branch formation from olfactory bulb output neurons. Cell. 2002;109(2):217-28.
70. Schwanzel-Fukuda M, Bick D, Pfaff DW. Luteinizing hormone-releasing hormone (LHRH)-expressing cells do not migrate normally in an inherited hypogonadal (Kallmann) syndrome. Brain research Molecular brain research. 1989;6(4):311-26.
71. Cariboni A, Pimpinelli F, Colamarino S, Zaninetti R, Piccolella M, Rumio C, et al. The product of X-linked Kallmann's syndrome gene (KAL1) affects the migratory activity of gonadotropin-releasing hormone (GnRH)-producing neurons. Human molecular genetics. 2004;13(22):2781-91.
72. Yanicostas C, Herbomel E, Dipietromaria A, Soussi-Yanicostas N. Anosmin-1a is required for fasciculation and terminal targeting of olfactory sensory neuron axons in the zebrafish olfactory system. Molecular and cellular endocrinology. 2009;312(1-2):53-60.
73. Kramer PR, Wray S. Novel gene expressed in nasal region influences outgrowth of olfactory axons and migration of luteinizing hormone-releasing hormone (LHRH) neurons. Genes Dev. 2000;14(14):1824-34.
74. Palevitch O, Abraham E, Borodovsky N, Levkowitz G, Zohar Y, Gothilf Y. Nasal embryonic LHRH factor plays a role in the developmental migration and projection of gonadotropin-releasing hormone 3 neurons in zebrafish. Developmental dynamics : an official publication of the American Association of Anatomists. 2009;238(1):66-75.
75. Cadman SM, Kim SH, Hu Y, Gonzalez-Martinez D, Bouloux PM. Molecular pathogenesis of Kallmann's syndrome. Hormone research. 2007;67(5):231-42.

76. Gill JC, Tsai PS. Expression of a dominant negative FGF receptor in developing GNRH1 neurons disrupts axon outgrowth and targeting to the median eminence. Biology of reproduction. 2006;74(3):463-72.
77. Tornberg J, Sykiotis GP, Keefe K, Plummer L, Hoang X, Hall JE, et al. Heparan sulfate 6-O-sulfotransferase 1, a gene involved in extracellular sugar modifications, is mutated in patients with idiopathic hypogonadotrophic hypogonadism. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(28):11524-9.
78. Ojeda SR, Lomniczi A, Mastronardi C, Heger S, Roth C, Parent AS, et al. Minireview: the neuroendocrine regulation of puberty: is the time ripe for a systems biology approach? Endocrinology. 2006;147(3):1166-74.
79. Colledge WH, Mei H, d'Anglemont de Tassigny X. Mouse models to study the central regulation of puberty. Molecular and cellular endocrinology. 2010;324(1-2):12-20.
80. Ojeda SR, Prevot V, Heger S, Lomniczi A, Dziedzic B, Mungenast A. Glia-to-neuron signaling and the neuroendocrine control of female puberty. Ann Med. 2003;35(4):244-55.
81. Erecinska M, Silver IA. Metabolism and role of glutamate in mammalian brain. Prog Neurobiol. 1990;35(4):245-96.
82. Plant TM, Gay VL, Marshall GR, Arslan M. Puberty in monkeys is triggered by chemical stimulation of the hypothalamus. Proceedings of the National Academy of Sciences of the United States of America. 1989;86(7):2506-10.
83. de Roux N, Genin E, Carel JC, Matsuda F, Chaussain JL, Milgrom E. Hypogonadotropic hypogonadism due to loss of function of the KiSS1-derived peptide receptor GPR54. Proceedings of the National Academy of Sciences of the United States of America. 2003;100(19):10972-6.
84. Seminara SB, Messager S, Chatzidaki EE, Thresher RR, Acierno JS, Jr., Shagoury JK, et al. The GPR54 gene as a regulator of puberty. N Engl J Med. 2003;349(17):1614-27.
85. Topaloglu AK, Tello JA, Kotan LD, Ozbek MN, Yilmaz MB, Erdogan S, et al. Inactivating KISS1 mutation and hypogonadotropic hypogonadism. N Engl J Med. 2012;366(7):629-35.
86. Silveira LG, Noel SD, Silveira-Neto AP, Abreu AP, Brito VN, Santos MG, et al. Mutations of the KISS1 gene in disorders of puberty. The Journal of clinical endocrinology and metabolism. 2010;95(5):2276-80.

87. Navarro VM, Fernandez-Fernandez R, Castellano JM, Roa J, Mayen A, Barreiro ML, et al. Advanced vaginal opening and precocious activation of the reproductive axis by KiSS-1 peptide, the endogenous ligand of GPR54. J Physiol. 2004;561(Pt 2):379-86.
88. Tena-Sempere M. Kisspeptin signaling in the brain: recent developments and future challenges. Molecular and cellular endocrinology. 2010;314(2):164-9.
89. Pinilla L, Aguilar E, Dieguez C, Millar RP, Tena-Sempere M. Kisspeptins and reproduction: physiological roles and regulatory mechanisms. Physiol Rev. 2012;92(3):1235-316.
90. Topaloglu AK, Reimann F, Guclu M, Yalin AS, Kotan LD, Porter KM, et al. TAC3 and TACR3 mutations in familial hypogonadotropic hypogonadism reveal a key role for Neurokinin B in the central control of reproduction. Nature genetics. 2009;41(3):354-8.
91. Guran T, Tolhurst G, Bereket A, Rocha N, Porter K, Turan S, et al. Hypogonadotropic hypogonadism due to a novel missense mutation in the first extracellular loop of the neurokinin B receptor. The Journal of clinical endocrinology and metabolism. 2009;94(10):3633-9.
92. Rance NE. Menopause and the human hypothalamus: evidence for the role of kisspeptin/neurokinin B neurons in the regulation of estrogen negative feedback. Peptides. 2009;30(1):111-22.
93. Sandoval-Guzman T, Rance NE. Central injection of senktide, an NK3 receptor agonist, or neuropeptide Y inhibits LH secretion and induces different patterns of Fos expression in the rat hypothalamus. Brain research. 2004;1026(2):307-12.
94. Kung TT, Crawley Y, Jones H, Luo B, Gilchrest H, Greenfeder S, et al. Tachykinin NK3-receptor deficiency does not inhibit pulmonary eosinophilia in allergic mice. Pharmacological research : the official journal of the Italian Pharmacological Society. 2004;50(6):611-5.
95. Ramaswamy S, Seminara SB, Plant TM. Evidence from the agonadal juvenile male rhesus monkey (Macaca mulatta) for the view that the action of neurokinin B to trigger gonadotropin-releasing hormone release is upstream from the kisspeptin receptor. Neuroendocrinology. 2011;94(3):237-45.
96. Mitsushima D, Hei DL, Terasawa E. gamma-Aminobutyric acid is an inhibitory neurotransmitter restricting the release of luteinizing hormone-releasing hormone before

the onset of puberty. Proceedings of the National Academy of Sciences of the United States of America. 1994;91(1):395-9.
97. Ducret E, Anderson GM, Herbison AE. RFamide-related peptide-3, a mammalian gonadotropin-inhibitory hormone ortholog, regulates gonadotropin-releasing hormone neuron firing in the mouse. Endocrinology. 2009;150(6):2799-804.
98. Ojeda SR, Lomniczi A, Sandau US. Glial-gonadotrophin hormone (GnRH) neurone interactions in the median eminence and the control of GnRH secretion. Journal of neuroendocrinology. 2008;20(6):732-42.
99. Voigt P, Ma YJ, Gonzalez D, Fahrenbach WH, Wetsel WC, Berg-von der Emde K, et al. Neural and glial-mediated effects of growth factors acting via tyrosine kinase receptors on luteinizing hormone-releasing hormone neurons. Endocrinology. 1996;137(6):2593-605.
100. Prevot V, Rio C, Cho GJ, Lomniczi A, Heger S, Neville CM, et al. Normal female sexual development requires neuregulin-erbB receptor signaling in hypothalamic astrocytes. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2003;23(1):230-9.
101. Garcia-Segura LM, Melcangi RC. Steroids and glial cell function. Glia. 2006;54(6):485-98.
102. Ojeda SR, Dubay C, Lomniczi A, Kaidar G, Matagne V, Sandau US, et al. Gene networks and the neuroendocrine regulation of puberty. Molecular and cellular endocrinology. 2010;324(1-2):3-11.
103. Ojeda SR, Hill J, Hill DF, Costa ME, Tapia V, Cornea A, et al. The Oct-2 POU domain gene in the neuroendocrine brain: a transcriptional regulator of mammalian puberty. Endocrinology. 1999;140(8):3774-89.
104. Lee BJ, Cho GJ, Norgren RB, Jr., Junier MP, Hill DF, Tapia V, et al. TTF-1, a homeodomain gene required for diencephalic morphogenesis, is postnatally expressed in the neuroendocrine brain in a developmentally regulated and cell-specific fashion. Mol Cell Neurosci. 2001;17(1):107-26.
105. Heger S, Mastronardi C, Dissen GA, Lomniczi A, Cabrera R, Roth CL, et al. Enhanced at puberty 1 (EAP1) is a new transcriptional regulator of the female neuroendocrine reproductive axis. The Journal of clinical investigation. 2007;117(8):2145-54.

106. Siler-Khodr TM, Khodr GS. Studies in human fetal endocrinology. I. Luteinizing hormone-releasing factor content of the hypothalamus. American journal of obstetrics and gynecology. 1978;130(7):795-800.
107. Huhtaniemi IT, Howard S, Dunkel L, Anderson RA. The Gonadal Axis: A Life Perspective. In: Pfaff DWaJ, M, editor. Hormones, Brain, and Behavior, 3rd edition. Vol 4. : Oxford: Academic Press; 2017. p. 3–58.
108. Waldhauser F, Weissenbacher G, Frisch H, Pollak A. Pulsatile secretion of gonadotropins in early infancy. Eur J Pediatr. 1981;137(1):71-4.
109. Dunkel L, Alfthan H, Stenman UH, Tapanainen P, Perheentupa J. Pulsatile secretion of LH and FSH in prepubertal and early pubertal boys revealed by ultrasensitive time-resolved immunofluorometric assays. Pediatric research. 1990;27(3):215-9.
110. Albertsson-Wikland K, Rosberg S, Lannering B, Dunkel L, Selstam G, Norjavaara E. Twenty-four-hour profiles of luteinizing hormone, follicle-stimulating hormone, testosterone, and estradiol levels: a semilongitudinal study throughout puberty in healthy boys. The Journal of clinical endocrinology and metabolism. 1997;82(2):541-9.
111. Chemes HE, Gottlieb SE, Pasqualini T, Domenichini E, Rivarola MA, Bergada C. Response to acute hCG stimulation and steroidogenic potential of Leydig cell fibroblastic precursors in humans. J Androl. 1985;6(2):102-12.
112. Rey RA, Campo SM, Bedecarras P, Nagle CA, Chemes HE. Is infancy a quiescent period of testicular development? Histological, morphometric, and functional study of the seminiferous tubules of the cebus monkey from birth to the end of puberty. The Journal of clinical endocrinology and metabolism. 1993;76(5):1325-31.
113. Brambilla DJ, Matsumoto AM, Araujo AB, McKinlay JB. The effect of diurnal variation on clinical measurement of serum testosterone and other sex hormone levels in men. The Journal of clinical endocrinology and metabolism. 2009;94(3):907-13.
114. Ankarberg-Lindgren C, Norjavaara E. Changes of diurnal rhythm and levels of total and free testosterone secretion from pre to late puberty in boys: testis size of 3 ml is a transition stage to puberty. European journal of endocrinology / European Federation of Endocrine Societies. 2004;151(6):747-57.
115. Andersson AM, Juul A, Petersen JH, Muller J, Groome NP, Skakkebaek NE. Serum inhibin B in healthy pubertal and adolescent boys: relation to age, stage of puberty, and

follicle-stimulating hormone, luteinizing hormone, testosterone, and estradiol levels. The Journal of clinical endocrinology and metabolism. 1997;82(12):3976-81.
116. Andersson AM, Toppari J, Haavisto AM, Petersen JH, Simell T, Simell O, et al. Longitudinal reproductive hormone profiles in infants: peak of inhibin B levels in infant boys exceeds levels in adult men. The Journal of clinical endocrinology and metabolism. 1998;83(2):675-81.
117. Raivio T, Perheentupa A, McNeilly AS, Groome NP, Anttila R, Siimes MA, et al. Biphasic increase in serum inhibin B during puberty: a longitudinal study of healthy Finnish boys. Pediatric research. 1998;44(4):552-6.
118. Byrd W, Bennett MJ, Carr BR, Dong Y, Wians F, Rainey W. Regulation of biologically active dimeric inhibin A and B from infancy to adulthood in the male. The Journal of clinical endocrinology and metabolism. 1998;83(8):2849-54.
119. Bergada I, Rojas G, Ropelato G, Ayuso S, Bergada C, Campo S. Sexual dimorphism in circulating monomeric and dimeric inhibins in normal boys and girls from birth to puberty. Clinical endocrinology. 1999;51(4):455-60.
120. Knorr D, Bidlingmaier F, Butenandt O, Fendel H. Plasma testosterone in male puberty. I. Physiology of plasma testosterone. Acta endocrinologica. 1974;75(1):181-94.
121. Widen E, Silventoinen K, Sovio U, Ripatti S, Cousminer DL, Hartikainen AL, et al. Pubertal timing and growth influences cardiometabolic risk factors in adult males and females. Diabetes care. 2012;35(4):850-6.
122. Ritte R, Lukanova A, Tjonneland A, Olsen A, Overvad K, Mesrine S, et al. Height, age at menarche and risk of hormone receptor positive and negative breast cancer: A cohort study. International journal of cancer Journal international du cancer. 2012.
123. He C, Zhang C, Hunter DJ, Hankinson SE, Buck Louis GM, Hediger ML, et al. Age at menarche and risk of type 2 diabetes: results from 2 large prospective cohort studies. American journal of epidemiology. 2010;171(3):334-44.
124. Kvale G, Heuch I, Nilssen S. Re: "Endometrial cancer and age at last delivery: evidence for an association". Am J Epidemiol. 1992;135(4):453-5.
125. Purdie DM, Green AC. Epidemiology of endometrial cancer. Best Pract Res Clin Obstet Gynaecol. 2001;15(3):341-54.

126. Elks CE, Loos RJ, Sharp SJ, Langenberg C, Ring SM, Timpson NJ, et al. Genetic markers of adult obesity risk are associated with greater early infancy weight gain and growth. PLoS Med. 2010;7(5):e1000284.
127. Lakshman R, Forouhi NG, Sharp SJ, Luben R, Bingham SA, Khaw KT, et al. Early age at menarche associated with cardiovascular disease and mortality. The Journal of clinical endocrinology and metabolism. 2009;94(12):4953-60.
128. Prentice P, Viner RM. Pubertal timing and adult obesity and cardiometabolic risk in women and men: a systematic review and meta-analysis. International journal of obesity. 2013;37(8):1036-43.
129. Ong KK, Northstone K, Wells JC, Rubin C, Ness AR, Golding J, et al. Earlier mother's age at menarche predicts rapid infancy growth and childhood obesity. PLoS medicine. 2007;4(4):e132.
130. Day FR, Elks CE, Murray A, Ong KK, Perry JR. Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank study. Sci Rep. 2015;5:11208.
131. Albanese A, Stanhope R. Does constitutional delayed puberty cause segmental disproportion and short stature? Eur J Pediatr. 1993;152(4):293-6.
132. Bierich JR, Nolte K, Drews K, Brugmann G. Constitutional delay of growth and adolescence. Results of short-term and long-term treatment with GH. Acta endocrinologica. 1992;127(5):392-6.
133. Crowne EC, Shalet SM, Wallace WH, Eminson DM, Price DA. Final height in girls with untreated constitutional delay in growth and puberty. Eur J Pediatr. 1991;150(10):708-12.
134. Crowne EC, Shalet SM, Wallace WH, Eminson DM, Price DA. Final height in boys with untreated constitutional delay in growth and puberty. Archives of disease in childhood. 1990;65(10):1109-12.
135. Moreira-Andres MN, Canizo FJ, de la Cruz FJ, Gomez-de la Camara A, Hawkins FG. Bone mineral status in prepubertal children with constitutional delay of growth and puberty. European journal of endocrinology / European Federation of Endocrine Societies. 1998;139(3):271-5.
136. Finkelstein JS, Neer RM, Biller BM, Crawford JD, Klibanski A. Osteopenia in men with a history of delayed puberty. N Engl J Med. 1992;326(9):600-4.

137. Crowne EC, Shalet SM. Management of constitutional delay in growth and puberty. Trends in endocrinology and metabolism: TEM. 1990;1(5):239-42.
138. Kaltiala-Heino R, Kosunen E, Rimpela M. Pubertal timing, sexual behaviour and self-reported depression in middle adolescence. J Adolesc. 2003;26(5):531-45.
139. Sedlmeyer IL, Palmert MR. Delayed puberty: analysis of a large case series from an academic center. The Journal of clinical endocrinology and metabolism. 2002;87(4):1613-20.
140. LaFranchi S, Hanna CE, Mandel SH. Constitutional delay of growth: expected versus final adult height. Pediatrics. 1991;87(1):82-7.
141. Albanese A, Stanhope R. Predictive factors in the determination of final height in boys with constitutional delay of growth and puberty. The Journal of pediatrics. 1995;126(4):545-50.
142. Wehkalampi K, Vangonen K, Laine T, Dunkel L. Progressive reduction of relative height in childhood predicts adult stature below target height in boys with constitutional delay of growth and puberty. Hormone research. 2007;68(2):99-104.
143. Rensonnet C, Kanen F, Coremans C, Ernould C, Albert A, Bourguignon JP. Pubertal growth as a determinant of adult height in boys with constitutional delay of growth and puberty. Hormone research. 1999;51(5):223-9.
144. Volta C, Ghizzoni L, Buono T, Ferrari F, Virdis R, Bernasconi S. Final height in a group of untreated children with constitutional growth delay. Helv Paediatr Acta. 1988;43(3):171-6.
145. Bramswig JH, Fasse M, Holthoff ML, von Lengerke HJ, von Petrykowski W, Schellong G. Adult height in boys and girls with untreated short stature and constitutional delay of growth and puberty: accuracy of five different methods of height prediction. The Journal of pediatrics. 1990;117(6):886-91.
146. Arrigo T, Cisternino M, Luca De F, Saggese G, Messina MF, Pasquino AM, et al. Final height outcome in both untreated and testosterone-treated boys with constitutional delay of growth and puberty. Journal of pediatric endocrinology & metabolism : JPEM. 1996;9(5):511-7.
147. Sperlich M, Butenandt O, Schwarz HP. Final height and predicted height in boys with untreated constitutional growth delay. Eur J Pediatr. 1995;154(8):627-32.
148. Cools BL, Rooman R, Op De Beeck L, Du Caju MV. Boys with a simple delayed puberty reach their target height. Hormone research. 2008;70(4):209-14.

149. Dunkel L, Huhtaniemi I. Abnormal prolactin secretion in prepubertal boys with hypogonadotrophic hypogonadism--possible involvement in regulation of testicular steroidogenesis. International journal of andrology. 1985;8(5):385-92.
150. Zevenhuijzen H, Kelnar CJ, Crofton PM. Diagnostic utility of a low-dose gonadotropin-releasing hormone test in the context of puberty disorders. Hormone research. 2004;62(4):168-76.
151. Dunkel L, Perheentupa J, Virtanen M, Maenpaa J. Gonadotropin-releasing hormone test and human chorionic gonadotropin test in the diagnosis of gonadotropin deficiency in prepubertal boys. The Journal of pediatrics. 1985;107(3):388-92.
152. Segal TY, Mehta A, Anazodo A, Hindmarsh PC, Dattani MT. Role of gonadotropin-releasing hormone and human chorionic gonadotropin stimulation tests in differentiating patients with hypogonadotropic hypogonadism from those with constitutional delay of growth and puberty. J Clin Endocrinol Metab. 2009;94(3):780-5.
153. Degros V, Cortet-Rudelli C, Soudan B, Dewailly D. The human chorionic gonadotropin test is more powerful than the gonadotropin-releasing hormone agonist test to discriminate male isolated hypogonadotropic hypogonadism from constitutional delayed puberty. European journal of endocrinology / European Federation of Endocrine Societies. 2003;149(1):23-9.
154. Martin MM, Martin AL. Constitutional delayed puberty in males and hypogonadotropic hypogonadism: a reliable and cost-effective approach to differential diagnosis. Journal of pediatric endocrinology & metabolism : JPEM. 2005;18(9):909-16.
155. Demir A, Voutilainen R, Juul A, Dunkel L, Alfthan H, Skakkebaek NE, et al. Increase in first morning voided urinary luteinizing hormone levels precedes the physical onset of puberty. J Clin Endocrinol Metab. 1996;81(8):2963-7.
156. Coutant R, Biette-Demeneix E, Bouvattier C, Bouhours-Nouet N, Gatelais F, Dufresne S, et al. Baseline inhibin B and anti-Mullerian hormone measurements for diagnosis of hypogonadotropic hypogonadism (HH) in boys with delayed puberty. J Clin Endocrinol Metab. 2010;95(12):5225-32.
157. Rohayem J, Nieschlag E, Kliesch S, Zitzmann M. Inhibin B, AMH, but not INSL3, IGF1 or DHEAS support differentiation between constitutional delay of growth and puberty and hypogonadotropic hypogonadism. Andrology. 2015;3(5):882-7.

158. Silveira LF, Latronico AC. Approach to the patient with hypogonadotropic hypogonadism. The Journal of clinical endocrinology and metabolism. 2013;98(5):1781-8.
159. Laitinen EM, Tommiska J, Sane T, Vaaralahti K, Toppari J, Raivio T. Reversible congenital hypogonadotropic hypogonadism in patients with CHD7, FGFR1 or GNRHR mutations. PloS one. 2012;7(6):e39450.
160. Raivio T, Falardeau J, Dwyer A, Quinton R, Hayes FJ, Hughes VA, et al. Reversal of idiopathic hypogonadotropic hypogonadism. N Engl J Med. 2007;357(9):863-73.
161. Massin N, Pecheux C, Eloit C, Bensimon JL, Galey J, Kuttenn F, et al. X chromosome-linked Kallmann syndrome: clinical heterogeneity in three siblings carrying an intragenic deletion of the KAL-1 gene. The Journal of clinical endocrinology and metabolism. 2003;88(5):2003-8.
162. Pedersen-White JR, Chorich LP, Bick DP, Sherins RJ, Layman LC. The prevalence of intragenic deletions in patients with idiopathic hypogonadotropic hypogonadism and Kallmann syndrome. Molecular human reproduction. 2008;14(6):367-70.
163. Albuisson J, Pecheux C, Carel JC, Lacombe D, Leheup B, Lapuzina P, et al. Kallmann syndrome: 14 novel mutations in KAL1 and FGFR1 (KAL2). Human mutation. 2005;25(1):98-9.
164. Franco B, Guioli S, Pragliola A, Incerti B, Bardoni B, Tonlorenzi R, et al. A gene deleted in Kallmann's syndrome shares homology with neural cell adhesion and axonal path-finding molecules. Nature. 1991;353(6344):529-36.
165. Harrington J, Palmert MR. Clinical review: Distinguishing constitutional delay of growth and puberty from isolated hypogonadotropic hypogonadism: critical appraisal of available diagnostic tests. The Journal of clinical endocrinology and metabolism. 2012;97(9):3056-67.
166. Juul A, Aksglaede L, Bay K, Grigor KM, Skakkebaek NE. Klinefelter syndrome: the forgotten syndrome: basic and clinical questions posed to an international group of scientists. Acta paediatrica. 2011;100(6):791-2.
167. Blevins CH, Wilson ME. Klinefelter's syndrome. Bmj. 2012;345:e7558.
168. Saenger P, Wikland KA, Conway GS, Davenport M, Gravholt CH, Hintz R, et al. Recommendations for the diagnosis and management of Turner syndrome. The Journal of clinical endocrinology and metabolism. 2001;86(7):3061-9.

169. Cox L, Liu JH. Primary ovarian insufficiency: an update. International journal of women's health. 2014;6:235-43.
170. Wallace WH, Shalet SM, Crowne EC, Morris-Jones PH, Gattamaneni HR. Ovarian failure following abdominal irradiation in childhood: natural history and prognosis. Clin Oncol (R Coll Radiol). 1989;1(2):75-9.
171. Ahonen P, Myllarniemi S, Sipila I, Perheentupa J. Clinical variation of autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) in a series of 68 patients. N Engl J Med. 1990;322(26):1829-36.
172. Aittomaki K, Lucena JL, Pakarinen P, Sistonen P, Tapanainen J, Gromoll J, et al. Mutation in the follicle-stimulating hormone receptor gene causes hereditary hypergonadotropic ovarian failure. Cell. 1995;82(6):959-68.
173. Partsch CJ, Heger S, Sippell WG. Management and outcome of central precocious puberty. Clinical endocrinology. 2002;56(2):129-48.
174. Biro FM, Galvez MP, Greenspan LC, Succop PA, Vangeepuram N, Pinney SM, et al. Pubertal assessment method and baseline characteristics in a mixed longitudinal study of girls. Pediatrics. 2010;126(3):e583-90.
175. Kim SH, Huh K, Won S, Lee KW, Park MJ. A Significant Increase in the Incidence of Central Precocious Puberty among Korean Girls from 2004 to 2010. PloS one. 2015;10(11):e0141844.
176. Herman-Giddens ME, Slora EJ, Wasserman RC, Bourdony CJ, Bhapkar MV, Koch GG, et al. Secondary sexual characteristics and menses in young girls seen in office practice: a study from the Pediatric Research in Office Settings network. Pediatrics. 1997;99(4):505-12.
177. Kaplowitz PB, Oberfield SE. Reexamination of the age limit for defining when puberty is precocious in girls in the United States: implications for evaluation and treatment. Drug and Therapeutics and Executive Committees of the Lawson Wilkins Pediatric Endocrine Society. Pediatrics. 1999;104(4 Pt 1):936-41.
178. Midyett LK, Moore WV, Jacobson JD. Are pubertal changes in girls before age 8 benign? Pediatrics. 2003;111(1):47-51.
179. Bridges NA, Christopher JA, Hindmarsh PC, Brook CG. Sexual precocity: sex incidence and aetiology. Archives of disease in childhood. 1994;70(2):116-8.

180. Teilmann G, Pedersen CB, Jensen TK, Skakkebaek NE, Juul A. Prevalence and incidence of precocious pubertal development in Denmark: an epidemiologic study based on national registries. Pediatrics. 2005;116(6):1323-8.

181. Caronia LM MC, Welt CK, Sykiotis GP, Quinton R, Thambundit A, Avbelj M, Dhruvakumar S, Plummer L, Hughes VA, Seminara SB,Boepple PA, Sidis Y, Crowley WF, Martin KA, Hall JE, Pitteloud N. A Genetic Basis for Functional Hypothalamic Amenorrhea. N Engl J Med. 2011;364:215-25.

182. Alikasifoglu A, Vuralli D, Gonc EN, Ozon A, Kandemir N. Changing Etiological Trends in Male Precocious Puberty: Evaluation of 100 Cases with Central Precocious Puberty over the Last Decade. Hormone research in paediatrics. 2015;83(5):340-4.

183. Martin Diaz MJ, Soriano Guillen L, Munoz Calvo MT, Pozo Roman J, Argente Oliver J. [Central precocious puberty is associated with a high prevalence of organic disease]. An Pediatr (Barc). 2006;65(5):434-8.

184. Chemaitilly W, Merchant TE, Li Z, Barnes N, Armstrong GT, Ness KK, et al. Central Precocious Puberty following the Diagnosis and Treatment of Paediatric Cancer and Central Nervous System Tumours: Presentation and Long-term Outcomes. Clinical endocrinology. 2015.

185. de Vries L, Kauschansky A, Shohat M, Phillip M. Familial central precocious puberty suggests autosomal dominant inheritance. The Journal of clinical endocrinology and metabolism. 2004;89(4):1794-800.

186. Teles MG, Bianco SD, Brito VN, Trarbach EB, Kuohung W, Xu S, et al. A GPR54-activating mutation in a patient with central precocious puberty. N Engl J Med. 2008;358(7):709-15.

187. Abreu AP, Dauber A, Macedo DB, Noel SD, Brito VN, Gill JC, et al. Central precocious puberty caused by mutations in the imprinted gene MKRN3. N Engl J Med. 2013;368(26):2467-75.

188. Simon D, Ba I, Mekhail N, Ecosse E, Paulsen A, Zenaty D, et al. Mutations in the maternally imprinted gene MKRN3 are common in familial central precocious puberty. European journal of endocrinology / European Federation of Endocrine Societies. 2016;174(1):1-8.

189. Macedo DB, Abreu AP, Reis AC, Montenegro LR, Dauber A, Beneduzzi D, et al. Central precocious puberty that appears to be sporadic caused by paternally inherited

mutations in the imprinted gene makorin ring finger 3. The Journal of clinical endocrinology and metabolism. 2014;99(6):E1097-103.
190. Lee HS, Jin HS, Shim YS, Jeong HR, Kwon E, Choi V, et al. Low Frequency of MKRN3 Mutations in Central Precocious Puberty Among Korean Girls. Horm Metab Res. 2015.
191. Soriano-Guillen L, Mitchell V, Carel JC, Barbet P, Roger M, Lahlou N. Activating mutations in the luteinizing hormone receptor gene: a human model of non-follicle-stimulating hormone-dependent inhibin production and germ cell maturation. The Journal of clinical endocrinology and metabolism. 2006;91(8):3041-7.
192. Stratakis CA, Vottero A, Brodie A, Kirschner LS, DeAtkine D, Lu Q, et al. The aromatase excess syndrome is associated with feminization of both sexes and autosomal dominant transmission of aberrant P450 aromatase gene transcription. The Journal of clinical endocrinology and metabolism. 1998;83(4):1348-57.
193. Low LC, Wang Q. Gonadotropin independent precocious puberty. Journal of pediatric endocrinology & metabolism : JPEM. 1998;11(4):497-507.
194. Armengaud JB, Charkaluk ML, Trivin C, Tardy V, Breart G, Brauner R, et al. Precocious pubarche: distinguishing late-onset congenital adrenal hyperplasia from premature adrenarche. The Journal of clinical endocrinology and metabolism. 2009;94(8):2835-40.
195. Schwindinger WF, Francomano CA, Levine MA. Identification of a mutation in the gene encoding the alpha subunit of the stimulatory G protein of adenylyl cyclase in McCune-Albright syndrome. Proceedings of the National Academy of Sciences of the United States of America. 1992;89(11):5152-6.
196. Elks CE, Perry JR, Sulem P, Chasman DI, Franceschini N, He C, et al. Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. Nature genetics. 2010;42(12):1077-85.
197. Ong KK, Elks CE, Li S, Zhao JH, Luan J, Andersen LB, et al. Genetic variation in LIN28B is associated with the timing of puberty. Nature genetics. 2009;41(6):729-33.
198. Perry JR, Day F, Elks CE, Sulem P, Thompson DJ, Ferreira T, et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. Nature. 2014;514(7520):92-7.
199. Perry JR, Stolk L, Franceschini N, Lunetta KL, Zhai G, McArdle PF, et al. Meta-analysis of genome-wide association data identifies two loci influencing age at menarche. Nature genetics. 2009;41(6):648-50.

200. Tommiska J, Wehkalampi K, Vaaralahti K, Laitinen EM, Raivio T, Dunkel L. LIN28B in constitutional delay of growth and puberty. The Journal of clinical endocrinology and metabolism. 2010;95(6):3063-6.
201. Silveira-Neto AP, Leal LF, Emerman AB, Henderson KD, Piskounova E, Henderson BE, et al. Absence of functional LIN28B mutations in a large cohort of patients with idiopathic central precocious puberty. Hormone research in paediatrics. 2012;78(3):144-50.
202. Cousminer DL, Stergiakouli E, Berry DJ, Ang W, Groen-Blokhuis MM, Korner A, et al. Genome-wide association study of sexual maturation in males and females highlights a role for body mass and menarche loci in male puberty. Human molecular genetics. 2014;23(16):4452-64.
203. Day FR, Bulik-Sullivan B, Hinds DA, Finucane HK, Murabito JM, Tung JY, et al. Shared genetic aetiology of puberty timing between sexes and with health-related outcomes. Nat Commun. 2015;6:8842.
204. Kaplowitz PB, Slora EJ, Wasserman RC, Pedlow SE, Herman-Giddens ME. Earlier onset of puberty in girls: relation to increased body mass index and race. Pediatrics. 2001;108(2):347-53.
205. Yeo GS. The role of the FTO (Fat Mass and Obesity Related) locus in regulating body size and composition. Molecular and cellular endocrinology. 2014;397(1-2):34-41.
206. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marin C, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature. 2014;507(7492):371-5.
207. Ballabio A, Bardoni B, Carrozzo R, Andria G, Bick D, Campbell L, et al. Contiguous gene syndromes due to deletions in the distal short arm of the human X chromosome. Proceedings of the National Academy of Sciences of the United States of America. 1989;86(24):10001-5.
208. Bianco SD, Kaiser UB. The genetic and molecular basis of idiopathic hypogonadotropic hypogonadism. Nature reviews Endocrinology. 2009;5(10):569-76.
209. Miraoui H, Dwyer AA, Sykiotis GP, Plummer L, Chung W, Feng B, et al. Mutations in FGF17, IL17RD, DUSP6, SPRY4, and FLRT3 are identified in individuals with congenital hypogonadotropic hypogonadism. American journal of human genetics. 2013;92(5):725-43.
210. Silveira LF, Trarbach EB, Latronico AC. Genetics basis for GnRH-dependent pubertal disorders in humans. Molecular and cellular endocrinology. 2010;324(1-2):30-8.

211. Dode C, Levilliers J, Dupont JM, De Paepe A, Le Du N, Soussi-Yanicostas N, et al. Loss-of-function mutations in FGFR1 cause autosomal dominant Kallmann syndrome. Nature genetics. 2003;33(4):463-5.
212. Dode C, Teixeira L, Levilliers J, Fouveaut C, Bouchard P, Kottler ML, et al. Kallmann syndrome: mutations in the genes encoding prokineticin-2 and prokineticin receptor-2. PLoS genetics. 2006;2(10):e175.
213. Herbison AE. Genetics of puberty. Hormone research. 2007;68 Suppl 5:75-9.
214. Pitteloud N, Quinton R, Pearce S, Raivio T, Acierno J, Dwyer A, et al. Digenic mutations account for variable phenotypes in idiopathic hypogonadotropic hypogonadism. The Journal of clinical investigation. 2007;117(2):457-63.
215. Sykiotis GP, Plummer L, Hughes VA, Au M, Durrani S, Nayak-Young S, et al. Oligogenic basis of isolated gonadotropin-releasing hormone deficiency. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(34):15140-4.
216. Chevrier L, Guimiot F, de Roux N. GnRH receptor mutations in isolated gonadotropic deficiency. Molecular and cellular endocrinology. 2011;346(1-2):21-8.
217. Themmen APN, Huhtaniemi IT. Mutations of gonadotropins and gonadotropin receptors: elucidating the physiology and pathophysiology of pituitary-gonadal function. Endocrine reviews. 2000;21(5):551-83.
218. Layman LC, Lee EJ, Peak DB, Namnoum AB, Vu KV, van Lingen BL, et al. Delayed puberty and hypogonadism caused by mutations in the follicle-stimulating hormone beta-subunit gene. N Engl J Med. 1997;337(9):607-11.
219. Sedlmeyer IL. Pedigree Analysis of Constitutional Delay of Growth and Maturation: Determination of Familial Aggregation and Inheritance Patterns. Journal of Clinical Endocrinology & Metabolism. 2002;87(12):5581-6.
220. Wehkalampi K, Widen E, Laine T, Palotie A, Dunkel L. Association of the timing of puberty with a chromosome 2 locus. The Journal of clinical endocrinology and metabolism. 2008;93(12):4833-9.
221. Cousminer DL, Leinonen JT, Sarin AP, Chheda H, Surakka I, Wehkalampi K, et al. Targeted resequencing of the pericentromere of chromosome 2 linked to constitutional delay of growth and puberty. PloS one. 2015;10(6):e0128524.
222. Lin L, Conway GS, Hill NR, Dattani MT, Hindmarsh PC, Achermann JC. A homozygous R262Q mutation in the gonadotropin-releasing hormone receptor presenting as

constitutional delay of growth and puberty with subsequent borderline oligospermia. The Journal of clinical endocrinology and metabolism. 2006;91(12):5117-21.
223. Vaaralahti K, Wehkalampi K, Tommiska J, Laitinen EM, Dunkel L, Raivio T. The role of gene defects underlying isolated hypogonadotropic hypogonadism in patients with constitutional delay of growth and puberty. Fertility and sterility. 2011;95(8):2756-8.
224. Pugliese-Pires PN, Fortin JP, Arthur T, Latronico AC, Mendonca BB, Villares SM, et al. Novel inactivating mutations in the GH secretagogue receptor gene in patients with constitutional delay of growth and puberty. European journal of endocrinology / European Federation of Endocrine Societies. 2011;165(2):233-41.
225. Tusset C, Noel SD, Trarbach EB, Silveira LF, Jorge AA, Brito VN, et al. Mutational analysis of TAC3 and TACR3 genes in patients with idiopathic central pubertal disorders. Arquivos brasileiros de endocrinologia e metabologia. 2012;56(9):646-52.
226. Zhu J, Choa RE, Guo MH, Plummer L, Buck C, Palmert MR, et al. A Shared Genetic Basis for Self-Limited Delayed Puberty and Idiopathic Hypogonadotropic Hypogonadism. The Journal of clinical endocrinology and metabolism. 2015:jc20151080.
227. Pitteloud N, Meysing A, Quinton R, Acierno JS, Jr., Dwyer AA, Plummer L, et al. Mutations in fibroblast growth factor receptor 1 cause Kallmann syndrome with a wide spectrum of reproductive phenotypes. Molecular and cellular endocrinology. 2006;254-255:60-9.
228. Caligioni CS. Assessing reproductive status/stages in mice. Current protocols in neuroscience / editorial board, Jacqueline N Crawley [et al]. 2009;Appendix 4:Appendix 4I.
229. Bedell VM, Westcot SE, Ekker SC. Lessons from morpholino-based screening in zebrafish. Briefings in functional genomics. 2011;10(4):181-8.
230. Abraham E, Palevitch O, Gothilf Y, Zohar Y. The zebrafish as a model system for forebrain GnRH neuronal development. Gen Comp Endocrinol. 2009;164(2-3):151-60.
231. Chen W, Ge W. Gonad differentiation and puberty onset in the zebrafish: evidence for the dependence of puberty onset on body growth but not age in females. Molecular reproduction and development. 2013;80(5):384-92.
232. Belchetz PE, Plant TM, Nakai Y, Keogh EJ, Knobil E. Hypophysial responses to continuous and intermittent delivery of hypopthalamic gonadotropin-releasing hormone. Science. 1978;202(4368):631-3.

233. Plant TM. Neurobiological bases underlying the control of the onset of puberty in the rhesus monkey: a representative higher primate. Frontiers in neuroendocrinology. 2001;22(2):107-39.
234. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60.
235. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research. 2010;38(16):e164.
236. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nature methods. 2010;7(4):248-9.
237. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols. 2009;4(7):1073-81.
238. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. Behav Brain Res. 2001;125(1-2):279-84.
239. Saari A, Harju S, Makitie O, Saha MT, Dunkel L, Sankilampi U. Systematic growth monitoring for the early detection of celiac disease in children. JAMA Pediatr. 2015;169(3):e1525.
240. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. Nature protocols. 2009;4(3):363-71.
241. Jefferys BR, Kelley LA, Sternberg MJ. Protein folding requires crowd control in a simulated cell. Journal of molecular biology. 2010;397(5):1329-38.
242. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nature protocols. 2010;5(4):725-38.
243. Moult J, Fidelis K, Kryshtafovych A, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)--round IX. Proteins. 2011;79 Suppl 10:1-5.
244. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature methods. 2011;8(10):785-6.
245. Jo S, Vargyas M, Vasko-Szedlar J, Roux B, Im W. PBEQ-Solver for online visualization of electrostatic potential of biomolecules. Nucleic acids research. 2008;36(Web Server issue):W270-5.

246. Segev O, Samach A, Faerman A, Kalinski H, Beiman M, Gelfand A, et al. CMF608-a novel mechanical strain-induced bone-specific protein expressed in early osteochondroprogenitor cells. Bone. 2004;34(2):246-60.
247. Memi F, Abe P, Cariboni A, MacKay F, Parnavelas JG, Stumm R. CXC chemokine receptor 7 (CXCR7) affects the migration of GnRH neurons by regulating CXCL12 availability. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2013;33(44):17527-37.
248. Guasti L, Paul A, Laufer E, King P. Localization of Sonic hedgehog secreting and receiving cells in the developing and adult rat adrenal cortex. Molecular and cellular endocrinology. 2011;336(1-2):117-22.
249. Abraham E, Palevitch O, Ijiri S, Du SJ, Gothilf Y, Zohar Y. Early development of forebrain gonadotrophin-releasing hormone (GnRH) neurones and the role of GnRH as an autocrine migration factor. Journal of neuroendocrinology. 2008;20(3):394-405.
250. Jones DT, Cozzetto D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics. 2015;31(6):857-63.
251. Meyre D, Proulx K, Kawagoe-Takaki H, Vatin V, Gutierrez-Aguilar R, Lyon D, et al. Prevalence of loss-of-function FTO mutations in lean and obese individuals. Diabetes. 2010;59(1):311-8.
252. Ma M, Harding HP, O'Rahilly S, Ron D, Yeo GS. Kinetic analysis of FTO (fat mass and obesity-associated) reveals that it is unlikely to function as a sensor for 2-oxoglutarate. The Biochemical journal. 2012;444(2):183-7.
253. Sun Y, Bak B, Schoenmakers N, van Trotsenburg AS, Oostdijk W, Voshol P, et al. Loss-of-function mutations in IGSF1 cause an X-linked syndrome of central hypothyroidism and testicular enlargement. Nature genetics. 2012;44(12):1375-81.
254. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome research. 2010;20(1):110-21.
255. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. Genome research. 2009;19(9):1553-61.
256. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nature methods. 2014;11(4):361-2.

257. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Human mutation. 2013;34(1):57-65.
258. Cooper GM, Stone EA, Asimenos G, Program NCS, Green ED, Batzoglou S, et al. Distribution and intensity of constraint in mammalian genomic sequence. Genome research. 2005;15(7):901-13.
259. Kaplowitz PB. Link between body fat and the timing of puberty. Pediatrics. 2008;121 Suppl 3:S208-17.
260. Homma S, Shimada T, Hikake T, Yaginuma H. Expression pattern of LRR and Ig domain-containing protein (LRRIG protein) in the early mouse embryo. Gene expression patterns : GEP. 2009;9(1):1-26.
261. Radovick S, Wray S, Lee E, Nicols DK, Nakayama Y, Weintraub BD, et al. Migratory arrest of gonadotropin-releasing hormone neurons in transgenic mice. Proceedings of the National Academy of Sciences of the United States of America. 1991;88(8):3402-6.
262. Wetsel WC. Immortalized hypothalamic luteinizing hormone-releasing hormone (LHRH) neurons: a new tool for dissecting the molecular and cellular basis of LHRH physiology. Cellular and molecular neurobiology. 1995;15(1):43-78.
263. Gore AC, Roberts JL. Regulation of gonadotropin-releasing hormone gene expression in vivo and in vitro. Frontiers in neuroendocrinology. 1997;18(2):209-45.
264. Pimpinelli F, Redaelli E, Restano-Cassulini R, Curia G, Giacobini P, Cariboni A, et al. Depolarization differentially affects the secretory and migratory properties of two cell lines of immortalized luteinizing hormone-releasing hormone (LHRH) neurons. The European journal of neuroscience. 2003;18(6):1410-8.
265. Giacobini P, Giampietro C, Fioretto M, Maggi R, Cariboni A, Perroteau I, et al. Hepatocyte growth factor/scatter factor facilitates migration of GN-11 immortalized LHRH neurons. Endocrinology. 2002;143(9):3306-15.
266. Maggi R, Pimpinelli F, Molteni L, Milani M, Martini L, Piva F. Immortalized luteinizing hormone-releasing hormone neurons show a different migratory activity in vitro. Endocrinology. 2000;141(6):2105-12.
267. Toba Y, Tiong JD, Ma Q, Wray S. CXCR4/SDF-1 system modulates development of GnRH-1 neurons and the olfactory system. Developmental neurobiology. 2008;68(4):487-503.

268. Palevitch O, Abraham E, Borodovsky N, Levkowitz G, Zohar Y, Gothilf Y. Cxcl12a-Cxcr4b signaling is important for proper development of the forebrain GnRH system in zebrafish. Gen Comp Endocrinol. 2010;165(2):262-8.
269. Howard SR, Guasti L, Ruiz-Babot G, Mancini A, David A, Storr HL, et al. IGSF10 mutations dysregulate gonadotropin-releasing hormone neuronal migration resulting in delayed puberty. EMBO Mol Med. 2016.
270. Kjaer I, Fischer Hansen B. The human vomeronasal organ: prenatal developmental stages and distribution of luteinizing hormone-releasing hormone. Eur J Oral Sci. 1996;104(1):34-40.
271. Kastenhuber E, Kern U, Bonkowsky JL, Chien CB, Driever W, Schweitzer J. Netrin-DCC, Robo-Slit, and heparan sulfate proteoglycans coordinate lateral positioning of longitudinal dopaminergic diencephalospinal axons. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2009;29(28):8914-26.
272. Inatani M, Irie F, Plump AS, Tessier-Lavigne M, Yamaguchi Y. Mammalian brain morphogenesis and midline axon guidance require heparan sulfate. Science. 2003;302(5647):1044-6.
273. Clegg JM, Conway CD, Howe KM, Price DJ, Mason JO, Turnbull JE, et al. Heparan sulfotransferases Hs6st1 and Hs2st keep Erk in check for mouse corpus callosum development. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2014;34(6):2389-401.
274. Conway CD, Howe KM, Nettleton NK, Price DJ, Mason JO, Pratt T. Heparan sulfate sugar modifications mediate the functions of slits and other factors needed for mouse forebrain commissure development. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2011;31(6):1955-70.
275. Anower EKMF, Habuchi H, Nagai N, Habuchi O, Yokochi T, Kimata K. Heparan sulfate 6-O-sulfotransferase isoform-dependent regulatory effects of heparin on the activities of various proteases in mast cells and the biosynthesis of 6-O-sulfated heparin. The Journal of biological chemistry. 2013;288(6):3705-17.
276. Habuchi H, Nagai N, Sugaya N, Atsumi F, Stevens RL, Kimata K. Mice deficient in heparan sulfate 6-O-sulfotransferase-1 exhibit defective heparan sulfate biosynthesis, abnormal placentation, and late embryonic lethality. The Journal of biological chemistry. 2007;282(21):15578-88.

277. Loo BM, Salmivirta M. Heparin/Heparan sulfate domains in binding and signaling of fibroblast growth factor 8b. The Journal of biological chemistry. 2002;277(36):32616-23.
278. Hu Y, Guimond SE, Travers P, Cadman S, Hohenester E, Turnbull JE, et al. Novel mechanisms of fibroblast growth factor receptor 1 regulation by extracellular matrix protein anosmin-1. The Journal of biological chemistry. 2009;284(43):29905-20.
279. Villanueva C, de Roux N. FGFR1 mutations in Kallmann syndrome. Front Horm Res. 2010;39:51-61.
280. Conway CD, Price DJ, Pratt T, Mason JO. Analysis of axon guidance defects at the optic chiasm in heparan sulphate sulphotransferase compound mutant mice. J Anat. 2011;219(6):734-42.
281. Fischer J, Koch L, Emmerling C, Vierkotten J, Peters T, Bruning JC, et al. Inactivation of the Fto gene protects from obesity. Nature. 2009;458(7240):894-8.
282. Cole TJ, Flegal KM, Nicholls D, Jackson AA. Body mass index cut offs to define thinness in children and adolescents: international survey. Bmj. 2007;335(7612):194.
283. Nelson JF, Karelus K, Felicio LS, Johnson TE. Genetic influences on the timing of puberty in mice. Biology of reproduction. 1990;42(4):649-55.
284. Gill MS, Hall CM, Tillmann V, Clayton PE. Constitutional delay in growth and puberty (CDGP) is associated with hypoleptinaemia. Clinical endocrinology. 1999;50(6):721-6.
285. Church C, Moir L, McMurray F, Girard C, Banks GT, Teboul L, et al. Overexpression of Fto leads to increased food intake and results in obesity. Nature genetics. 2010;42(12):1086-92.
286. Speakman JR. The 'Fat Mass and Obesity Related' (FTO) gene: Mechanisms of Impact on Obesity and Energy Balance. Curr Obes Rep. 2015;4(1):73-91.
287. Gerken T, Girard CA, Tung YC, Webby CJ, Saudek V, Hewitson KS, et al. The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. Science. 2007;318(5855):1469-72.
288. McTaggart JS, Lee S, Iberl M, Church C, Cox RD, Ashcroft FM. FTO is expressed in neurones throughout the brain and its expression is unaltered by fasting. PloS one. 2011;6(11):e27968.
289. Joustra SD, Schoenmakers N, Persani L, Campi I, Bonomi M, Radetti G, et al. The IGSF1 deficiency syndrome: characteristics of male and female patients. The Journal of clinical endocrinology and metabolism. 2013;98(12):4942-52.

290. Joustra SD, Wehkalampi K, Oostdijk W, Biermasz NR, Howard S, Silander TL, et al. IGSF1 variants in boys with familial delayed puberty. Eur J Pediatr. 2015;174(5):687-92.
291. Goede J, Hack WW, Sijstermans K, van der Voort-Doedens LM, Van der Ploeg T, Meij-de Vries A, et al. Normative values for testicular volume measured by ultrasonography in a normal population from infancy to adolescence. Hormone research in paediatrics. 2011;76(1):56-64.
292. Leighton PA, Mitchell KJ, Goodrich LV, Lu X, Pinson K, Scherz P, et al. Defining brain wiring patterns and mechanisms through gene trapping in mice. Nature. 2001;410(6825):174-9.
293. Renaud J, Chedotal A. Time-lapse analysis of tangential migration in Sema6A and PlexinA2 knockouts. Mol Cell Neurosci. 2014;63:49-59.
294. Cariboni A, Davidson K, Rakic S, Maggi R, Parnavelas JG, Ruhrberg C. Defective gonadotropin-releasing hormone neuron migration in mice lacking SEMA3A signalling through NRP1 and NRP2: implications for the aetiology of hypogonadotropic hypogonadism. Human molecular genetics. 2011;20(2):336-44.
295. Hanchate NK, Giacobini P, Lhuillier P, Parkash J, Espy C, Fouveaut C, et al. SEMA3A, a gene involved in axonal pathfinding, is mutated in patients with Kallmann syndrome. PLoS genetics. 2012;8(8):e1002896.
296. Qu X, Wei H, Zhai Y, Que H, Chen Q, Tang F, et al. Identification, characterization, and functional study of the two novel human members of the semaphorin gene family. The Journal of biological chemistry. 2002;277(38):35574-85.
297. Garrett AM, Jucius TJ, Sigaud LP, Tang FL, Xiong WC, Ackerman SL, et al. Analysis of Expression Pattern and Genetic Deletion of Netrin5 in the Developing Mouse. Front Mol Neurosci. 2016;9:3.
298. Glinka A, Dolde C, Kirsch N, Huang YL, Kazanskaya O, Ingelfinger D, et al. LGR4 and LGR5 are R-spondin receptors mediating Wnt/beta-catenin and Wnt/PCP signalling. EMBO Rep. 2011;12(10):1055-61.
299. Styrkarsdottir U, Thorleifsson G, Sulem P, Gudbjartsson DF, Sigurdsson A, Jonasdottir A, et al. Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. Nature. 2013;497(7450):517-20.

300. Van Schoore G, Mendive F, Pochet R, Vassart G. Expression pattern of the orphan receptor LGR4/GPR48 gene in the mouse. Histochemistry and cell biology. 2005;124(1):35-50.

APPENDIX 1 – Hypogonadotropic Hypogonadism ('HH') GENE LIST

Gene
CHD7
DAX1
FGF8
FGFR1
GNRH1
GNRHR
HESX1
HS6ST1
KAL1
KISS1
KISS1R
LEP
LEPR
LHX3
LHX4
NELF
PC1
PROK2
PROKR2
PROP1
SEMA3A
SF1
TAC3
TAC3R
WDR11

25 'HH genes', mutations in which have been identified or implicated in the published literature as causal in patients with GnRH deficiency.

APPENDIX 2 – URLs

1000 Genomes, http://browser.1000genomes.org

Annovar, http:// annovar.openbioinformatics.org/

BioEdit, http://www.mbio.ncsu.edu/BioEdit/

Broad institute SNAP tool, http://www.broadinstitute.org/mpg/snap/ldsearch.php

dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/

Ensembl variant effect predictor, http:// www.ensembl.org/info/docs/tools/vep/;

ExAC Browser, http://exac.broadinstitute.org/

FATHMM, http://fathmm.biocompute.org.uk/

GATK resources, https://www.broadinstitute.org/gatk/guide/article?id=1247

Genego MetaCore, https://portal.genego.com/;

Ingenuity Variant Analysis, http:// www.qiagen.com/ingenuity

LRT, http://annovar.openbioinformatics.org/en/latest/user-guide/filter/

Mutation taster, http://www.mutationtaster.org/

NHLBI Exome Sequencing Project (ESP) Exome Variant Server, http://evs.gs.washington.edu/EVS/

Picard, http://broadinstitute.github.io/picard/

PolyPhen-2 v.2.2.2, http://genetics.bwh.harvard.edu/pph2/

SIFT v.1.03, http://sift.bii.a-star.edu.sg/

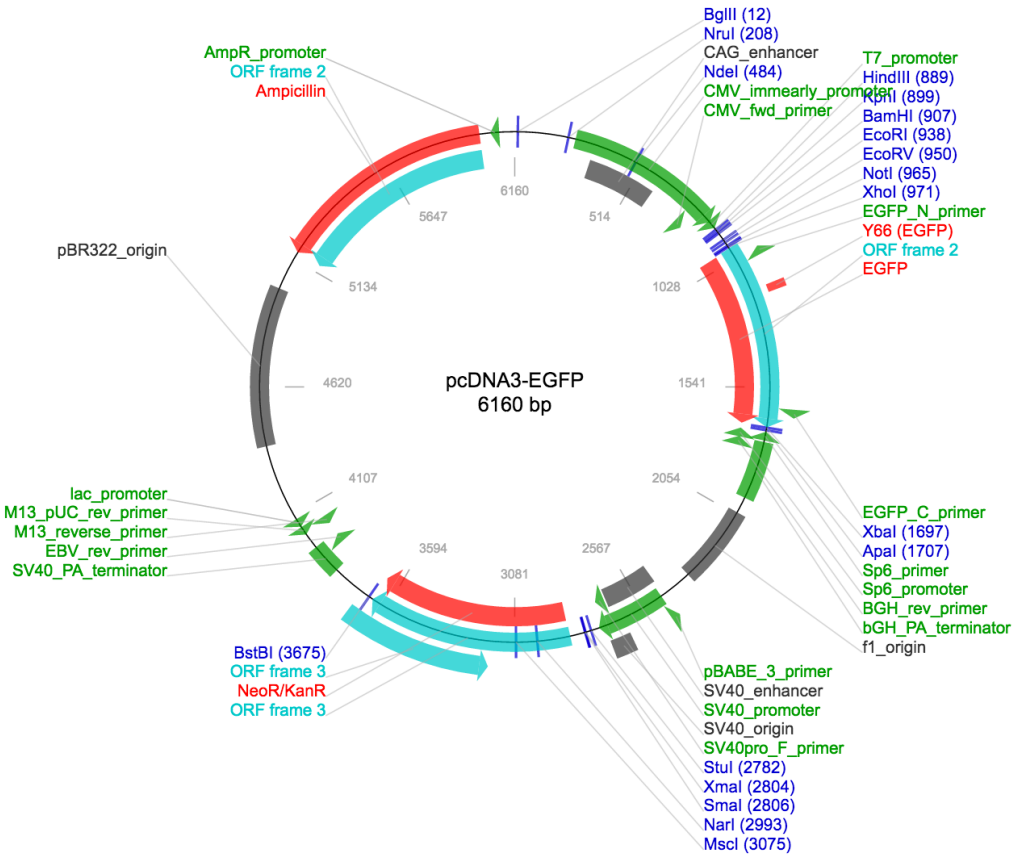
Sequencing Initiative Suomi, sisu.fimm.fi.

UniProt, http://www.uniprot.org/

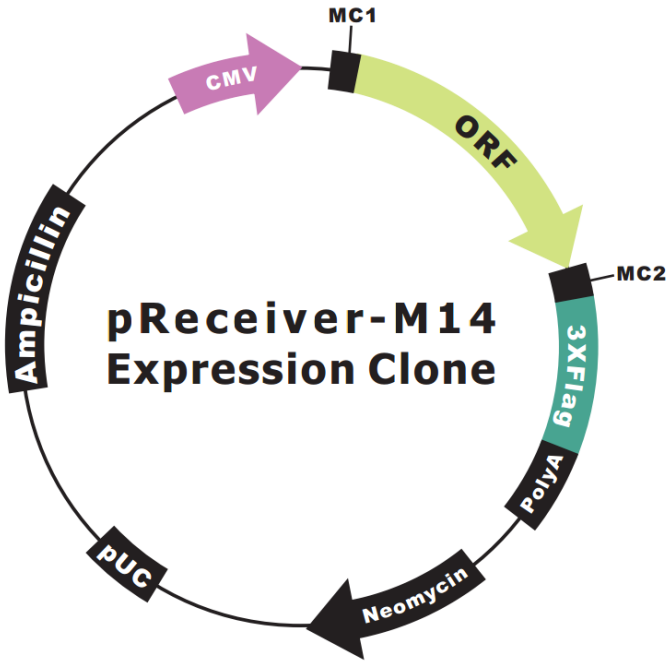
APPENDIX 3 – Vector maps

pcDNA3-EGFP vector used for *IGSF10* cloning (Addgene)

Vector Element	Utility
pcDNA3	Vector backbone – for mammalian expression
EGFP	Enhanced GFP insert for visual tracking of transduction and expression
NeoR	Neomycin (G418) selectable markers to permit antibiotic selection pressure and propagation of stable integrants
CMV promotor	Human cytomegalovirus promoter drives strong transgene expression
BGH	Bovine Growth Hormone polyadenylation signal and transcription termination sequence for enhanced mRNA stability
SV40 origin	SV40 origin for episomal replication and simple vector rescue in cell lines expressing the large T antigen
AmpR	Amplicillin resistance gene and pUC origin for selection and maintenance in bacterial cells



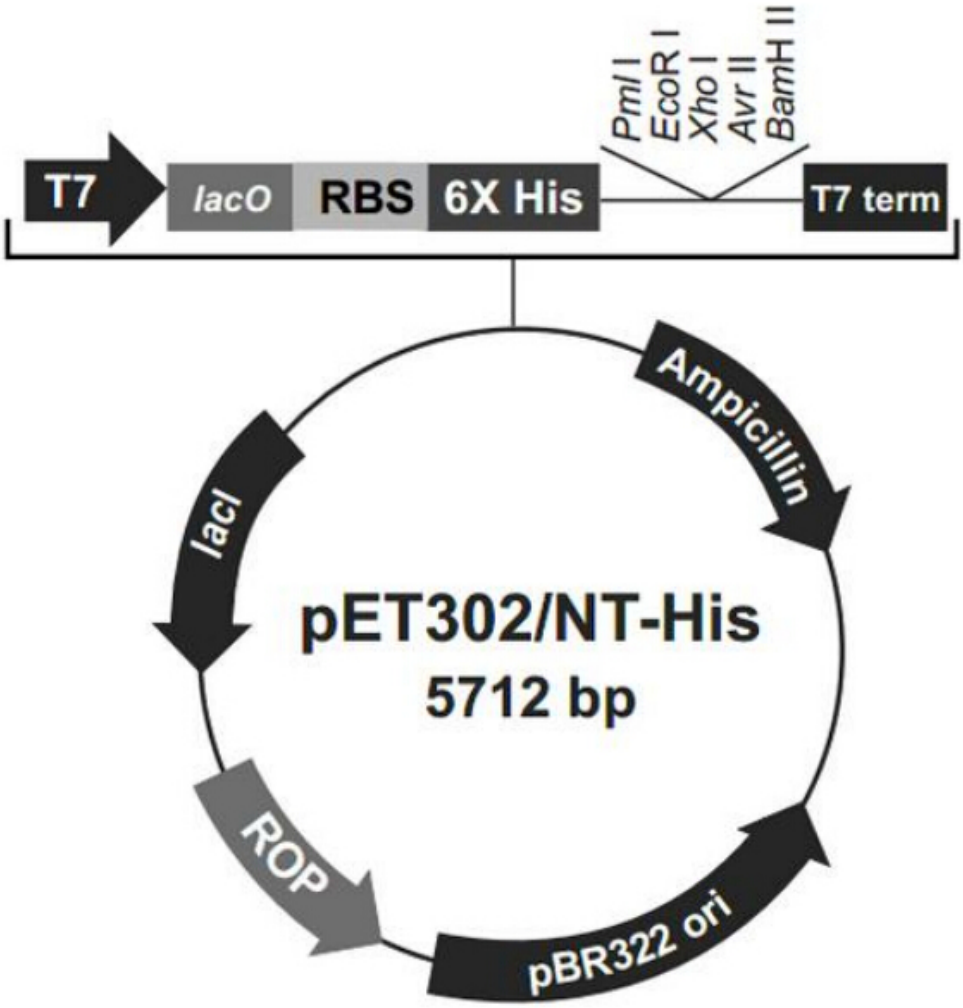
pReceiverM14a-WTh Vector used for *HS6ST1* cloning (Genecopoeia)



Vector Features

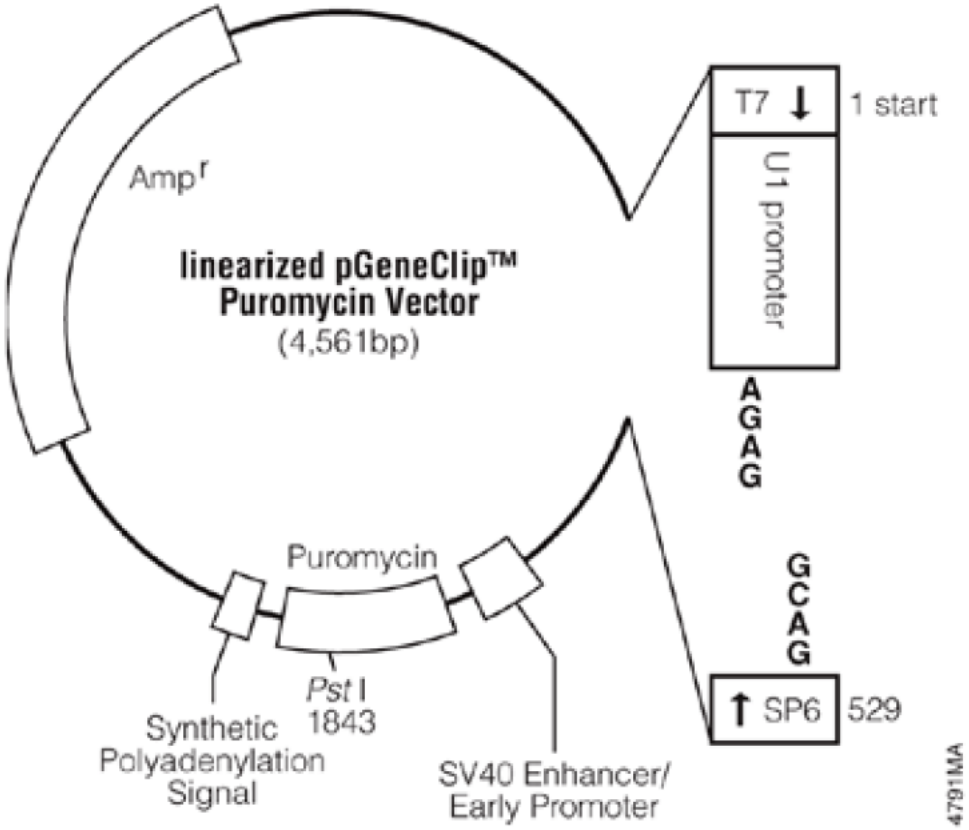
Promoter	CMV
Host Cell	Mammalian
Bacterial selection antibiotic	Ampicillin
Mammalian selection marker	Yes
Tag	3XFlag

pET302/NT-His vector used for *FTO* cloning (Invitrogen)



Vector Element	Utility
pET	Vector backbone – for mammalian expression
T7 promotor	Promoter drives strong transgene expression
AmpR	Amplicillin resistance gene for selection and maintenance in bacterial cells
Tag	His (6x)
Inducing agent	IPTG

SureSilencing shRNA plasmid pGeneClip™ Puromycin Vector used for *Igsf10* knockdown



Vector elements with sequence reference points: (Base pairs 4561)

T7 RNA polymerase transcription initiation site	1
U1 promoter	46-438
10bp spacer	439-448
U1 termination sequence	449-465
SP6 RNA polymerase promoter (-17 to +3)	527-546
SP6 RNA polymerase promoter primer binding site	529-547
Binding region of pUC/M13 reverse sequencing primer	564-585
SV40 early enhancer/promoter	798-1216
SV40 minimum origin of replication	1114-1179
Puromycine-N-acetyltransferase coding region	1239-1838
Synthetic poly(A) signal	1883-1931
Beta-lactamase (Ampr) coding region	2883-3743
Binding region of pUC/M13 forward sequencing primer	4495-4518
T7 RNA polymerase promoter (-17 to +3)	4545-3

APPENDIX 4 - GENERAL BUFFERS AND SOLUTIONS

dNTPs	Sequencing grade dNTPs [100mM] (Promega) were combined in a 1:1:1:1 ratio (dATP:dCTP:dGTP:dTTP) in dH_2O to produce a 10mM stock solution dNTP mix which was stored at -20C
IP/ Cell lysis buffer	50 mM Tris HCl pH 7.4, 500 mM NaCl, 1 mM EDTA, 0.5% Triton X-100
LB Agar	10g Agar (#A1296, Sigma), 10g LB Broth (# L3022, Sigma) in 500ml dH_2O
LB Broth	10g LB Broth in 500ml dH_2O
MOPS running buffer [1x]	50 ml MOPS SDS Running buffer [20x] (#NP0001-02, Invitrogen) in 1L dH_2O
PBS	5 tab Phosphate Buffer Saline (#P4417, Sigma) in 1L dH_2O, 0.01% Tween
PBS Tween	5 tab Phosphate Buffer Saline (#P4417, Sigma) in 1L dH_2O
Transfer buffer [1x]	2.42g Tris, 9g glycine, 800 ml dH2O, 200ml methanol
3M Sodium Acetate	Dissolve 40.8g sodium acetate in 80 ml H2O. Adjust pH to 5.2 with glacial acetic acid; make final volume up to 100ml and autoclave

APPENDIX 5 -
COMMERCIAL ASSAYS USED AND COMPOSITION OF REAGENTS

illustra Nucleon Genomic DNA Extraction Kit (GE Healthcare, Buckinghamshire, UK):

Reagent A (lysis agent): 10 mM Tris-HC1; 320 M sucrose; 5 mM MgCl2; 1%(v/v) Triton X-100; pH 8.0 (pH adjusted using 40%(w/v) NaOH)
Reagent B: (precise composition is not disclosed by the company)
5 M Sodium percholate
Nucleon resin: includes potassium hydroxide (precise composition is not disclosed by the company)

QIAprep Spin Miniprep Kit (Qiagen, Crawley, UK):

Buffer P1 (resuspension buffer): 50 mM Tris·Cl; 10 mM EDTA; 100 µg/ml RNase A
Buffer P2 (lysis buffer): 200 mM NaOH; 1% SDS (w/v)
Buffer N3 (neutralisation buffer): 4.2 M Guanidinium hydrochloride, 0.9M potassium acetate, pH 4.8
Buffer PB (wash buffer): 5 M Guanidinium hydrochloride; 30% isopropanol
Buffer PE (wash buffer): 10 mM Tris-HCl pH 7.5; 80% ethanol
Buffer EB (elution buffer): 10 mM Tris-Cl, pH 8.5
Other components of the kit: QIAprep Spin Columns, Collection Tubes (2 ml)

Qiagen HI SPEED Plasmid Midi Kit (Qiagen, Crawley, UK):

Buffer P1 (resuspension buffer): 50 mM Tris·Cl; 10 mM EDTA; 100 µg/ml RNase A
Buffer P2 (lysis buffer): 200 mM NaOH; 1% SDS (w/v)
Buffer P3 (neutralisation buffer): 3.0 M potassium acetate, pH 5.5
Buffer QBT (equilibration buffer): 750 mM NaCl; 50 mM MOPS, pH 7.0; 15% isopropanol (v/v); 0.15% Triton® X-100 (v/v)
Buffer QC (wash buffer): 1.0 M NaCl; 50 mM MOPS, pH 7.0; 15% isopropanol (v/v)
Buffer QF (elution buffer): 1.25 M NaCl; 50 mM Tris·Cl, pH 8.5; 15% isopropanol (v/v)
Other components of the kit: QIAfilter Cartridge; QIAGEN-tip

Qiagen RNeasy mini kit (Qiagen, Crawley, UK):

Buffer RLT (lysis buffer): contains a guanidine salt *(precise composition is not disclosed by the company)*
Buffer RW1 (stringent wash buffer): contains a guanidine salt, as well as ethanol *(precise composition is not disclosed by the company)*
Buffer RPE (mild wash buffer): Ethanol containing *(precise composition is not disclosed by the company)*
Other components of the kit: RNeasy mini spin columns (with silica membrane for RNA binding); Collection tubes (1.5; 2 ml); RNase-free water

Promega cDNA synthesis (M-MLV Reverse Transcriptase) (Promega, Southampton, UK):

Random primers: hexadeoxynucleotides, 500µl/ml
Reverse transcriptase enzyme (MMLV-RT enzyme)
Reverse transcriptase buffer (50mM Tris-HCl (pH8.3), 75 mM KCl, 3 mM MgCl2 and 10 mM DTT; 5x MMLV RT-buffer)

dNTPs (deoxynucleotidetriphosphates ; consisting of 10 mM dATP, 10 mM dCTP, 10 mM dGTP, 10 mM dTTP)
RNase inhibitor (RNasin)

Qiagen QIAquick Gel extraction kit (Qiagen, Crawley, UK):

Buffer QG (solubilisation and binding buffer, with pH indicator): 5.5 M guanidine thiocyanate, 20 M Tris HCl pH 6.6
Buffer PE (wash buffer): 10 mM Tris-HCl pH 7.5; 80% ethanol
Other components of the kit: QIAquick spin column; Collection tube (2 ml)

QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Stockport, UK)

PfuUltra High-Fidelity DNA polymerase (2.5 U/ µl)
10× reaction buffer
Dpn I restriction enzyme (10 U/µl)
Oligonucleotide control primer #1 [34-mer (100 ng/µl)] 5´ CCA TGA TTA CGC CAA GCG CGC AAT TAA CCC TCA C 3´
Oligonucleotide control primer #2 [34-mer (100 ng/µl)] 5´ GTG AGG GTT AAT TGC GCG CTT GGC GTA ATC ATG G 3´
pWhitescript 4.5-kb control plasmid (5 ng/ µl)
dNTP mix
XL1-Blue supercompetent cells
pUC18 control plasmid (0.1 ng/µl in TE buffer)

Universal Sulphotransferase Activity Kit (R&D systems, UK)

CHO cell-expressed Recombinant Mouse IMPAD1
PAP
Phosphatase Buffer 3
Phosphate Standard
Malachite Green Reagents A & B

APPENDIX 6 – PRIZES AND PRESENTATIONS

Prizes:

Highly commended in the Young Scientist competition, William Harvey Research Institute 30th Anniversary Conference, QMUL

June 2016

Endocrine Society Outstanding Abstract Award and Travel Award, American Endocrine Society April 2016

1st Oral Plenary Prize, Academy of Medical Sciences Spring Meeting for Clinician Scientists February 2016

1st Prize for Best Basic Science Oral Communication, British Endocrine Society November 2015

1st Prize for Best Oral Communication, British Society of Paediatric Endocrinology and Diabetes November 2015

Henning Andersen Award (Best Basic Abstract), European Society of Paediatric Endocrinology September 2015

Oral Communications:

Howard SR Searching for Genetic Gold at the End of the Puberty Rainbow, *April 2017, Endocrine Society Annual meeting (ENDO), Orlando, USA* – Invited Speaker

Howard SR, Guasti L, Ruiz-Babot G, Mancini A, David A, Storr HL, Metherell LA, Sternberg MJE, Cabrera CP, Warren HR, Barnes MR, Quinton R, de Roux N, Young J, Guiochon-Mantel A, Wehkalampi K, André V, Gothilf Y, Cariboni A, Dunkel L. Role of IGSF10 mutations in self-limited delayed puberty, *June 2016, William Harvey Research Institute 30th Anniversary Conference, London, UK*

Howard SR The Search for Disease-Causing Mutations in Self-Limited Delayed Puberty, *May 2016, European Congress of Endocrinology, Munich, Germany* – Invited Speaker

Howard SR, Guasti L, Ruiz-Babot G, Mancini A, David A, Storr HL, Metherell LA, Sternberg MJE, Cabrera CP, Warren HR, Barnes MR, Quinton R, de Roux N, Young J, Guiochon-Mantel A, Wehkalampi K, André V, Gothilf Y, Cariboni A, Dunkel L. 'Mutations in IGSF10 Cause Self-

Limited Delayed Puberty' and 'Searching for Genetic Gold at the End of the Puberty Rainbow', *April 2016, Endocrine Society Annual meeting (ENDO), Boston, USA*

Howard SR, Guasti L, Ruiz-Babot G, Mancini A, David A, Storr HL, Metherell LA, Sternberg MJE, Cabrera CP, Warren HR, Barnes MR, Quinton R, de Roux N, Young J, Guiochon-Mantel A, Wehkalampi K, André V, Gothilf Y, Cariboni A, Dunkel L. The Search for Disease-Causing Mutations in Self-Limited Delayed Puberty, *March 2016, COST Action BM1105 (GnRH network) Joint Scientific Meeting and Training School, Budapest, Hungary*

Howard SR, Guasti L, Ruiz-Babot G, Mancini A, David A, Storr HL, Metherell LA, Sternberg MJE, Cabrera CP, Warren HR, Barnes MR, Wehkalampi K, André V, Gothilf Y, Cariboni A, Dunkel L. Role of *IGSF10* mutations in self-limited delayed puberty, *February 2016, Academy of Medical Sciences Spring Meeting for Clinician Scientists in Training, London, UK*

Howard SR, Guasti L, Ruiz-Babot G, Mancini A, David A, Storr HL, Metherell LA, Sternberg MJE, Cabrera CP, Warren HR, Barnes MR, Wehkalampi K, André V, Gothilf Y, Cariboni A, Dunkel L. Mutations in *IGSF10* cause self-limited delayed puberty, via disturbance of GnRH neuronal migration*, November 2015, Wellcome Trust Joint Meeting of the Research Training Fellows and Clinical PhD Programme Appointees' meeting, London, UK*

Howard SR, Guasti L, Ruiz-Babot G, Mancini A, David A, Storr HL, Metherell LA, Sternberg MJE, Cabrera CP, Warren HR, Barnes MR, Wehkalampi K, André V, Gothilf Y, Cariboni A, Dunkel L. Mutations in *IGSF10* cause self-limited delayed puberty, via effects on GnRH neuronal migration *November 2015, British Society of Paediatric Endocrinology and Diabetes Annual Meeting, Sheffield, UK*

Howard SR, Poliandri A, Storr HL, Metherell LA, Cabrera CP, Warren HR, Barnes MR, Wehkalampi K, Guasti L, Dunkel L. Mutations in *HS6ST1* are causal in self-limited delayed puberty as well as idiopathic hypogonadotropic hypogonadism, *November 2015, Society for Endocrinology, British Endocrine Society annual meeting, Edinburgh, UK*

Howard SR, Guasti L, Ruiz-Babot G, Mancini A, David A, Storr HL, Metherell LA, Sternberg MJE, Cabrera CP, Warren HR, Barnes MR, Wehkalampi K, André V, Gothilf Y, Cariboni A, Dunkel L. Mutations in *IGSF10* cause self-limited delayed puberty, via disturbance of GnRH neuronal migration, *November 2015, Society for Endocrinology, British Endocrine Society annual meeting, Edinburgh, UK*

Howard SR, Poliandri A, Storr HL, Metherell LA, Cabrera CP, Warren HR, Barnes MR, Wehkalampi K, Guasti L, Dunkel L. A mutation in *HS6ST1* causes self-limited delayed puberty, *September 2015, European Society of Paediatric Endocrinology (ESPE) annual meeting, Barcelona, Spain*

Howard S, Guasti L, Ruiz-Babot G, Mancini A, David A, Storr HL, Metherell LA, Sternberg MJE, Cabrera CP, Warren HR, Barnes MR, Wehkalampi K, André V, Gothilf Y, Cariboni A,

Dunkel L. Mutations in *IGSF10* cause self-limited delayed puberty, via effects on GnRH neuronal migration, *September 2015, European Society of Paediatric Endocrinology (ESPE) annual meeting, Barcelona, Spain*

Howard S, Guasti L, Mancini A, David A, Storr HL, Metherell LA, Sternberg MJE, Cabrera CP, Warren HR, Barnes MR, Wehkalampi K, André V, Gothilf Y, Cariboni A, Dunkel L. Mutations in *IGSF10* Contribute to the Phenotype of Constitutional Delay of Growth and Puberty, via effects on GnRH neuronal migration, *April 2015, COST Action BM1105 (GnRH network) Joint Scientific Meeting and Training School, Prato, Italy*

Howard S, Guasti L, Barnes MR, Cabrera C, Metherell LA., Storr HL, Wehkalampi K and Dunkel L Novel Genes Affecting the Timing of Puberty. *Nov 2013. British Society of Paediatric Endocrinology and Diabetes Annual Meeting, Brighton, UK*

Howard S, Guasti L, Barnes MR, Cabrera C, Metherell LA., Storr HL, Wehkalampi K and Dunkel L Novel Genes Affecting the Timing of Puberty. *Oct 2013. William Harvey Research Institute Annual Research Review, London, UK*

Howard S, Guasti L, Barnes MR, Cabrera C, Metherell LA., Storr HL, Wehkalampi K and Dunkel L Investigating the Genetic Basis of Delayed Puberty. *Sept 2013. European Society of Paediatric Endocrinology Annual Meeting, Milan, Italy*

Poster Presentations:

Howard SR, Poliandri A, Storr HL, Metherell LA, Cabrera CP, Warren HR, Barnes MR, Wehkalampi K, Guasti L, Dunkel L Mutations in *HS6ST1* Cause Self-Limited Delayed Puberty, in addition to Idiopathic Hypogonadotropic Hypogonadism, *March 2016. COST Action BM1105 (GnRH network) Joint Scientific Meeting and Training School, Budapest, Hungary*

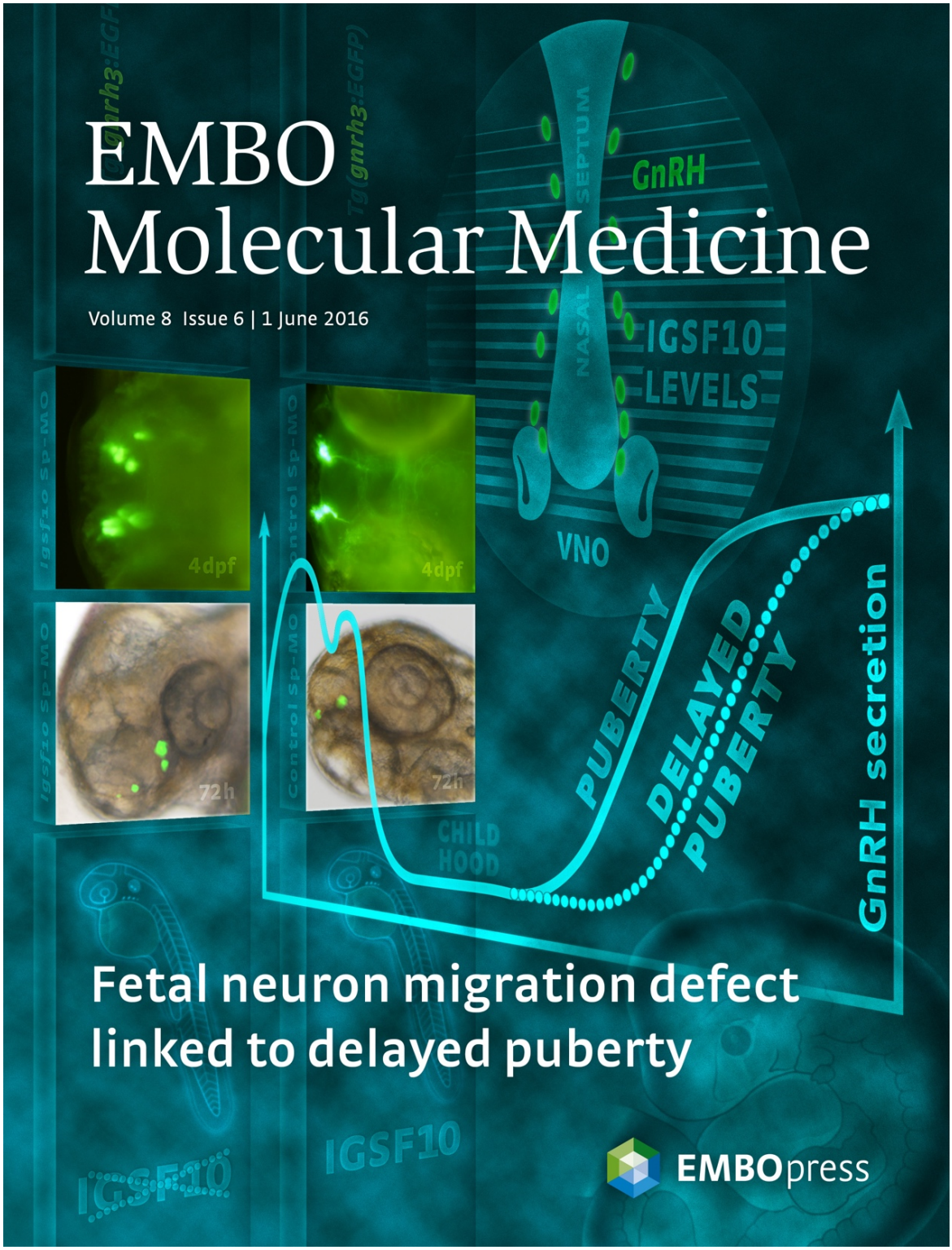
Howard SR, Poliandri A, Storr HL, Metherell LA, Cabrera CP, Warren HR, Barnes MR, Wehkalampi K, Guasti L, Dunkel L Mutations in *HS6ST1* Cause Self-Limited Delayed Puberty, in addition to Idiopathic Hypogonadotropic Hypogonadism, *Nov 2015. British Society of Paediatric Endocrinology and Diabetes Annual Meeting, Sheffield, UK*

Howard S, Guasti L, Storr HL, Metherell LM, Barnes M, Cabrera C, Warren H, Wehkalampi K, Cariboni A and Dunkel L Novel Gene Variants Influencing the Timing of Puberty, *April 2014 COST Action BM1105 (GnRH network) Joint Scientific Meeting and Training School, Berlin, Germany*

Howard S, Guasti L, Storr HL, Metherell LM, Barnes M, Cabrera C, Warren H, Wehkalampi K, Cariboni A and Dunkel L A Novel Gene Affecting the Timing of Puberty*, March 2014 The Society for Endocrinology British Endocrine Society Annual Meeting, Liverpool, UK*

Howard S, Guasti L, Barnes MR, Cabrera C, Metherell LA, Storr HL, Wehkalampi K and Dunkel L Novel Genes Affecting the Timing of Puberty. *Sept 2013. Wellcome Trust- Nature Genetics 'Genomics of Common Diseases' Joint Meeting, Oxford, UK*

Howard S, Barnes MR, Metherell LA., Storr HL, Wehkalampi K and Dunkel L Familial Constitutional Delay in Growth and Puberty (self-limited DP) is a condition with significant genetic heterogeneity and limited overlap with the timing of puberty in the general population. *March 2013 The Society for Endocrinology British Endocrine Society Annual Meeting, Harrogate, UK*


EMBO Molecular Medicine
Volume 8 Issue 6 | 1 June 2016
Tg(gnrh3:EGFP)
Igsf10 Sp-MO
Control Sp-MO
4dpf
72h
GnRH
NASAL SEPTUM
IGSF10 LEVELS
VNO
PUBERTY
DELAYED PUBERTY
CHILD HOOD
GnRH secretion
Fetal neuron migration defect linked to delayed puberty
IGSF10
EMBOpress

Published online: April 13, 2016

Research Article

SOURCE DATA | TRANSPARENT PROCESS | OPEN ACCESS

EMBO Molecular Medicine

IGSF10 mutations dysregulate gonadotropin-releasing hormone neuronal migration resulting in delayed puberty

Sasha R Howard[†,1], Leonardo Guasti[†,1], Gerard Ruiz-Babot[1], Alessandra Mancini[1], Alessia David[2], Helen L Storr[1], Lousie A Metherell[1], Michael JE Sternberg[2], Claudia P Cabrera[3,4], Helen R Warren[4,5], Michael R Barnes[3,4], Richard Quinton[6], Nicolas de Roux[7,8,9], Jacques Young[10,11,12,13], Anne Guiochon-Mantel[10,11,12], Karoliina Wehkalampi[14], Valentina André[15], Yoav Gothilf[16], Anna Cariboni[15,17] & Leo Dunkel[1,*]

Abstract

Early or late pubertal onset affects up to 5% of adolescents and is associated with adverse health and psychosocial outcomes. Self-limited delayed puberty (DP) segregates predominantly in an autosomal dominant pattern, but the underlying genetic background is unknown. Using exome and candidate gene sequencing, we have identified rare mutations in *IGSF10* in 6 unrelated families, which resulted in intracellular retention with failure in the secretion of mutant proteins. *IGSF10* mRNA was strongly expressed in embryonic nasal mesenchyme, during gonadotropin-releasing hormone (GnRH) neuronal migration to the hypothalamus. *IGSF10* knockdown caused a reduced migration of immature GnRH neurons *in vitro*, and perturbed migration and extension of GnRH neurons in a gnrh3:EGFP zebrafish model. Additionally, loss-of-function mutations in *IGSF10* were identified in hypothalamic amenorrhea patients. Our evidence strongly suggests that mutations in *IGSF10* cause DP in humans, and points to a common genetic basis for conditions of functional hypogonadotropic hypogonadism (HH). While dysregulation of GnRH neuronal migration is known to cause permanent HH, this is the first time that this has been demonstrated as a casual mechanism in DP.

Keywords delayed puberty; GnRH; hypothalamic amenorrhea; neuronal migration; puberty
Subject Categories Development & Differentiation; Genetics, Gene Therapy & Genetic Disease; Urogenital System
DOI 10.15252/emmm.201606250 | Received 26 January 2016 | Revised 1 March 2016 | Accepted 14 March 2016

Introduction

Puberty is the critical developmental stage during which reproductive capacity is attained. The onset of puberty is driven by the reactivation of the hypothalamic–pituitary–gonadal (HPG) axis after relative quiescence during childhood, with an increase in the pulsatile release of gonadotropin-releasing hormone (GnRH). While the timing of pubertal onset varies within and between different populations, it is a highly heritable trait, suggesting strong genetic

1 Centre for Endocrinology, William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK
2 Centre for Integrative Systems Biology and Bioinformatics, Department of Life Sciences, Imperial College London, London, UK
3 Centre for Translational Bioinformatics, William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK
4 NIHR Barts Cardiovascular Biomedical Research Unit, Queen Mary University of London, London, UK
5 Department of Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine, Queen Mary University of London, London, UK
6 Institute of Genetic Medicine University of Newcastle-upon-Tyne, Newcastle-upon-Tyne,UK
7 Unité Mixte de Recherche 1141, Institut National de la Santé et de la Recherche Médicale, Paris, France
8 Université Paris Diderot, Sorbonne Paris Cité, Hôpital Robert Debré, Paris, France
9 Laboratoire de Biochimie, Assistance Publique-Hôpitaux de Paris, Hôpital Robert Debré, Paris, France
10 Univ Paris-Sud, Le Kremlin Bicêtre, France
11 INSERM UMR-1185, Le Kremlin Bicêtre, France
12 Assistance Publique-Hôpitaux de Paris, Bicêtre Hospital, Le Kremlin-Bicêtre, France
13 Department of Reproductive Endocrinology, Bicêtre Hospital, Le Kremlin-Bicêtre, France
14 Children's Hospital, Helsinki University Hospital and University of Helsinki, Helsinki, Finland
15 Department of Pharmacological and Biomolecular Sciences, University of Milan, Milan, Italy
16 Department of Neurobiology, The George S. Wise Faculty of Life Sciences and Sagol School of Neuroscience, Tel-Aviv University, Tel Aviv, Israel
17 Institute of Ophthalmology, University College London (UCL), London, UK
*Corresponding author. Tel: +44 207 882 6235; Fax: +44 207 882 6197; E-mail: l.dunkel@qmul.ac.uk
†These authors contributed equally to this work

determinants (Wehkalampi *et al*, 2008b). Previous epidemiological studies estimate that 60–80% of the variation in pubertal onset is under genetic regulation (Parent *et al*, 2003; Gajdos *et al*, 2009; Morris *et al*, 2011). However, despite this strong heritability, little is known about the genetic control of human puberty (Palmert & Dunkel, 2012).

Abnormal pubertal timing affects up to 5% of adolescents and is associated with adverse health and psychosocial outcomes (He *et al*, 2010; Ritte *et al*, 2012; Widen *et al*, 2012; Day *et al*, 2015). Our lack of understanding of the factors that trigger pubertal onset is a barrier both to diagnosis and to the management of patients with pubertal disorders, and also hampers attempts to comprehend the population-wide trend toward an earlier age of pubertal onset in the developed world (DiVall & Radovick, 2008; Mouritsen *et al*, 2010).

Attempts to identify key genetic regulators of the timing of puberty have ranged from genome-wide association studies of age at menarche (Ong *et al*, 2009; Elks *et al*, 2010) to next-generation sequencing approaches. Together, these studies suggest that pubertal timing in the general population may be controlled by hundreds of genetic regulators, while loss-of-function mutations in one gene can produce the phenotypic features of complete GnRH deficiency. In patients with hypogonadotropic hypogonadism (HH), up to 30 separate genes resulting in severely delayed or absent puberty have been identified (Bianco & Kaiser, 2009; Gajdos *et al*, 2009). These genes control GnRH neuronal migration and differentiation, GnRH secretion, or its downstream pathways (Karges & de Roux, 2005; Beate *et al*, 2012). Evidence for digenic inheritance of HH, with synergistic effects of two gene defects together producing a more severe phenotype, has also been established (Pitteloud *et al*, 2007).

At the extreme end of the normal range of pubertal onset, self-limited delayed puberty (DP) is a common condition (Sedlmeyer, 2002a). Self-limited DP is defined as the absence of testicular enlargement in boys or breast development in girls at an age that is 2–2.5 standard deviations (SD) later than the population mean (Palmert & Dunkel, 2012). DP segregates within families, with the majority of families displaying an autosomal dominant pattern of inheritance (Sedlmeyer, 2002b; Wehkalampi *et al*, 2008b). Recently, variants in HH genes have been identified in some cases of self-limited DP (Zhu *et al*, 2015). However, in the majority of patients with DP, the neuroendocrine pathophysiology and its genetic regulation remain unclear. Our large, well-phenotyped cohort with self-limited DP from the relatively homogenous Finnish population provides invaluable familial data with which to investigate this question (Kristiansson *et al*, 2008; Wehkalampi *et al*, 2008b). We hypothesized that such families will be enriched for low-frequency, high- or moderate-effect alleles that are amenable to discovery through exome sequencing.

Results

Rare, potentially pathogenic variants in the *IGSF10* gene found in 10 families with DP

Initial whole exome sequencing performed in the 18 most extensive families from our cohort (111 individuals: a total of 76 individuals with DP, male = 53 and female = 23; and 35 controls, male = 13 and female = 22) identified 2,474,145 variants after quality control (Fig 1). Following filtering through our in-house pipeline to identify rare, predicted deleterious mutations, segregating with trait in an autosomal dominant inheritance pattern in multiple families and with potential biological relevance, 28 top candidate genes were identified. These 28 genes were then put forward for targeted resequencing in a further 42 families from the same cohort (178 individuals with DP and 110 controls, Fig 1), and the filtered results were analyzed by applying statistical thresholds for enrichment of rare, pathogenic variants in our cohort via rare variant burden testing with multiple comparison adjustment (Benjamini *et al*, 2001).

The candidate gene, *immunoglobulin superfamily member 10, IGSF10* (ENSG00000152580, gene identification number 285313), was identified after rare variant burden testing (adjusted *P*-value = 0.020) and screening of a further 100 controls from our cohort (Fig 1). Four genes had initially passed the $P < 0.025$ threshold after rare variant burden testing, and potentially pathogenic variants in these genes were further analyzed to determine their presence in controls from our cohort and for segregation with trait (Appendix Table S1 and Fig 1). Following this analysis, *IGSF10* was found to be the most promising candidate, with four potentially pathogenic variants in 10 probands from our cohort. The other 9 of 13 rare and potentially pathogenic variants that had been identified in *IGSF10* from targeted exome sequencing results were discarded in our post-sequencing analysis, as they were present in multiple controls from our cohort.

Four variants in *IGSF10* identified in 31 individuals from 10 families (NM_178822.4: c.467G>T (rs138756085) p.Arg156Leu, NM_178822.4: c.481G>A (rs114161831) p.Glu161Lys, NM_178822.4: c.6791A>G p.Glu2264Gly and NM_178822.4: c.7840G>A (rs112889898) p.Asp2614Asn) were found in ≤ 1 control subject (Table 1, Figs 2A and EV1).

Although three of the four variants were present in public databases, they were highly enriched in our cohort (Table 1). Analysis of self-limited DP families is complicated by the fact that this phenotype represents the tail of a normally distributed trait within the population, so it is anticipated that variants that govern the inheritance of this condition will also be present in the general population at a low level. Indeed, it is expected that up to 5% of the individuals sequenced in population databases will have abnormal pubertal timing, either early or delayed. Thus, the absence of these variants in population databases cannot be used as an exclusion criterion, and instead, a comparison of prevalence of such variants must be made to identify those that are enriched in patients compared to the ethnically matched general population.

All four *IGSF10* variants are heterozygous missense variants predicted to be deleterious, damaging, or possibly damaging by ≥ 3/5 prediction tools (Table 2). All variants affect amino acids that are highly conserved among homologues, as revealed by PhyloP or GERP score, and multiple sequence alignment (Table 2 and Appendix Fig S1).

Families with *IGSF10* variants display autosomal dominant inheritance and classical self-limited DP

Two N-terminal variants in *IGSF10* (p.Arg156Leu and p.Glu161Lys) were identified in 20 individuals from six families (Figs 2A and 3A). Perfect segregation with the expected autosomal dominant pattern of inheritance was seen in all but one individual

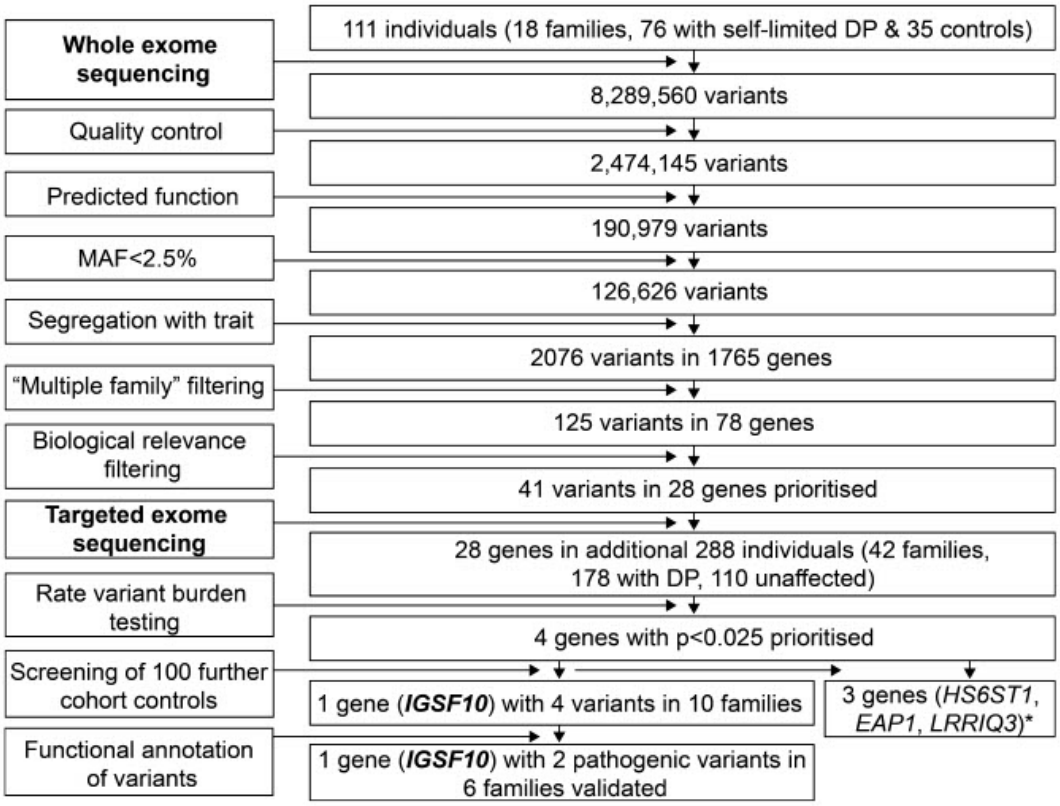


Figure 1. Flowchart of exome sequencing filtering outcomes.
Whole exome sequencing was initially performed on DNA extracted from the peripheral blood leukocytes of 111 individuals from the 18 most extensive families from our cohort (76 with DP and 35 controls). The exome sequences were aligned to the UCSC hg19 reference genome. Picard tools and the genome analysis toolkit were used to mark PCR duplicates, realign around indels, recalibrate quality scores, and call variants. Variants were analyzed further and filtered for potential causal variants using filters for quality control, predicted functional annotation, minor allele frequency (MAF), segregation with trait, variants in multiple families, and biological relevance (see Materials and Methods and Appendix Table S1 for further information on filtering criteria). Targeted exome sequencing using a Fluidigm array of 28 candidate genes identified post-filtering was then performed in a further 42 families from the same cohort (288 individuals, 178 with DP and 110 controls). Variants post-targeted resequencing were filtered using the same criteria as the whole exome sequencing data. Rare variant burden testing was performed for all genes selected for targeted resequencing, in order to rank candidate genes post-targeted resequencing. A multiple comparison adjustment was applied to the set of 28 *P*-values *post hoc* (Benjamini *et al*, 2001). Screening of 100 further cohort controls was via conventional Sanger sequencing. Functional annotation of the variants as described elsewhere in Materials and Methods. DP, delayed puberty. *data unpublished.

Table 1. Minor allele frequency of *IGSF10* variants in study population and control cohorts.

Nucleotide change	Amino acid change	Exon	MAF from DP patients (%) (n = 215)	MAF from controls (%) (n = 210)	MAF (%) Finnish/European/All
c.467G>T	p.Arg156Leu	3	2.8	0	0/0.5/0.4
c.481G>A	p.Glu161Lys	3	5.6	0.5	2.0/0.7/1.0
c.6791A>G	p.Glu2264Gly	6	0.5	0	not seen
c.7840G>A	p.Asp2614Asn	6	3.3	0	0/0.8/0.8

Minor allele frequency (MAF) data for the Finnish population were retrieved from The Sequencing Initiative Suomi (The SISu project) (http://www.sisuproject.fi/, release 3.0, accessed September 2015). European and other MAF data were retrieved from the ExAC Browser (Exome Aggregation Consortium (ExAC), Cambridge, MA: http://exac.broadinstitute.org, accessed September 2015).

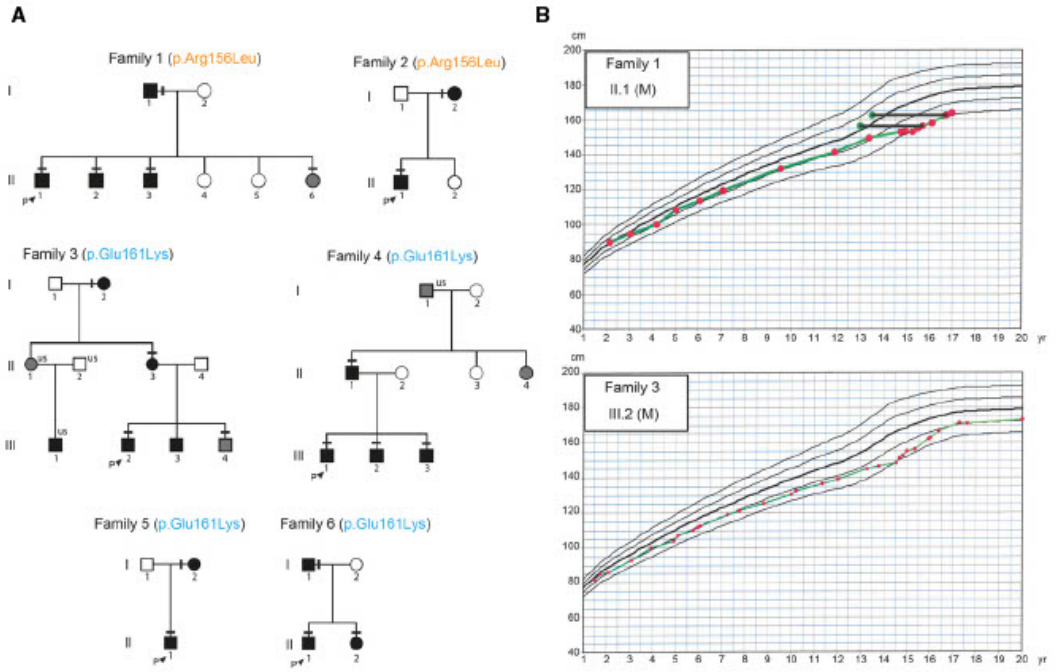


Figure 2. Pedigrees of the families with N-terminal *IGSF10* mutations with typical growth charts.

A Squares indicate male family members, and circles female family members. Black symbols represent clinically affected, gray symbols represent unknown phenotype, and clear symbols represent unaffected individuals. The arrow with "P" indicates the proband in each family and "us" indicates unsequenced due to the lack of DNA from that individual. The mutation in each family is given next to the family number; a horizontal black line above an individual's symbol indicates that they are heterozygous for that mutation as identified by either whole exome sequencing (family 3 and 4) or Fluidigm array (family 1, 2, 5, and 6), and verified by Sanger sequencing.

B Growth charts of 2 probands each showing typical growth patterns of self-limited DP, without compromised linear growth before puberty. Solid horizontal black lines connect green dots representing bone age to red dots at the equivalent chronological age.

Table 2. Prediction of *IGSF10* variants according to web-based prediction software programs and conservation across species.

AA Change	dbSNP137 ID	PhyloP (Pollard *et al*, 2010) Pred	SIFT (Kumar *et al*, 2009) Pred	PolyPhen-2 (Adzhubei *et al*, 2010) Pred	LRT (Chun & Fay, 2009) Pred	MutationTaster (Schwarz *et al*, 2014) Pred	FATHMM (Shihab *et al*, 2013) Pred	GERP (Cooper *et al*, 2005) ++
p.R156L	rs138756085	C	D	D	D	N	D	5.18
p.E161K	rs114161831	C	D	D	D	D	T	4.94
p.E2264G	n/a	C	D	P	N	D	T	3.71
p.D2614N	rs112889898	C	D	D	D	D	T	5.24

C, conserved; D, deleterious, disease causing or damaging; P, possibly damaging; N, neutral; T, tolerated.

(family3.III.3), who was found to have DP without carrying the variant. Of note given the known association between BMI and pubertal timing, this individual was very lean (weight 13% below median weight for height) at 13 years (Kaplowitz, 2008). The two C-terminal variants (p.Glu2264Gly and p.Asp2614Asn), identified in 11 individuals from four families, in contrast demonstrated incomplete penetrance in family 7 and a possible *de novo* mutation in family 10 (Fig EV1).

The affected individuals from these 10 families have classical clinical and biochemical features of "simple" DP, with delayed onset of Tanner stage 2 and delayed peak height velocity (Table 3). All probands had low gonadotropins with low or undetectable sex

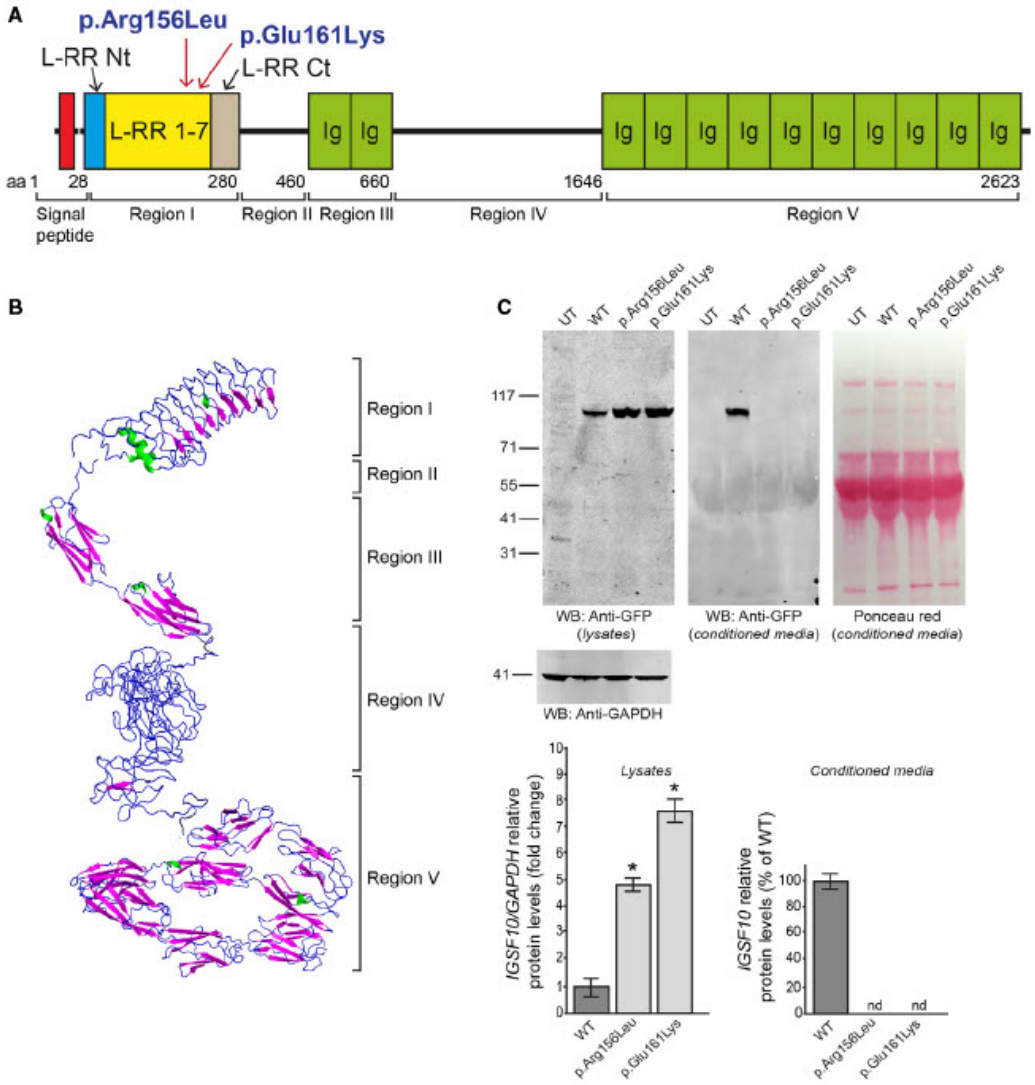


Figure 3. IGSF10 protein structure and position of N-terminal mutations.

A IGSF10 domains and N-terminal mutations identified in the study. Region I contains leucine-rich repeats (LRR) 1-7 flanked by a LRR N-terminal (LRR Nt) and C-terminal (LRR Ct) cap. Region II is structurally disordered. Region III contains two Ig-like domains (Ig). Region IV is structurally disordered. Region V contains 10 Ig-like domains (Ig).

B Protein tertiary structure as predicted by *in silico* analysis.

C Biological consequences of the 2 identified N-terminal mutations. Both WT and mutant N-terminal protein fragments (p.Arg156Leu and p.Glu161Lys) were expressed in HEK293 cells as demonstrated by Western blotting. The GFP-tagged protein products of both were not detected in the conditioned media of mammalian cells, as compared to wild type (WT), and appear to be retained in the intracellular compartment (mean ± SEM; $n = 3$). Ponceau red staining is shown to demonstrate equal protein loading for conditioned media. UT, untransfected negative control; nd, not detected; two-tailed *t*-test, $n = 3$ for each group, $^*P = 0.01094$ (WT vs. p.Arg156Leu) and $P = 0.04408$ (WT vs. p.Glu161Lys). It has to date not been possible to test cytoplasmic retention for the two C-terminal mutations due to difficulty expressing the full-length or C-terminal protein fragment in mammalian cells.

Source data are available online for this figure.

Table 3. Clinical and laboratory data of DP probands from each of the 10 families with potentially pathogenic mutations in *IGSF10*.

Case	1.II.1	2.II.1	3.III.2	4.III.1	5.II.1	6.II.1	7.III.5	8.III.2	9.II.4	10.II.1
Clinical data at 1st assessment:										
Sex	M	M	M	M	M	M	M	F	M	M
Age (years)	14.76	15.5	15.5	16.01	15.69	13.66	16.14	12.18	13.55	14.94
Bone age[a]	12.5	12.5	13.5	13.5	13.0	12.5	13	10.0	–	–
Tvol	2.0/2.0	2.0/2.0	4.0/4.0	2.0/2.0	2.0/2.0	2.0/2.0	2.0/2.0		2.0/2.0	2.0/2.0
Ph. G or B stage	1.1	1.1	1.2	2.1	1.1	1.1	1.1	1.1	1.1	1.1
BMI	19.1	18	19.4	24.9	17.9	22.8	18.6	14.5	14.5	15
LH[b] (IU/l) (0.1–0.6)[c]	0.3	0.1	0.4	0.1	0.2	0.1	0.1	0.1	0.15	0.15
FSH[b] (IU/l) (0.1–0.9)[c]	0.45	0.1	0.5	0.3	0.6	0.35	0.6	0.2	0.5	0.3
Testosterone (nmol/l) (0.1–1.0)[c]	0.3	0.22	0.4	0.55	0.3	0.22	0.2	–	0.4	0.4
estradiol (pg/ml) (< 8)[c]	–	–	–	–		–	–	< 5		
Inhibin B (pg/ml) (55–255)[c]	–	144	121	98	112	–	–		168	155
Age at:										
G2 or B2	15.2	15.6	15.5	16.5	16.10	15.21	16.5	13.94	15.21	15.4
Takeoff	14.76	15.81	15.6	16.11	15.59	15.6	16.5	13.94	15.5	15.91
PHV	15.6	16.2	17	17.3	16.2	16.18	17.3	14.68	16.1	16.4
Induction of puberty										
	Yes	No	No	Yes	Yes	Yes	Yes	No	No	No
Age at start	14.76			16.11	15.59	15.21	16.5			
Duration (months)	3			6	9	9	6			
Olfaction										
	Self-reported normal	Self-reported normal	Self-reported normal	Self-reported normal	Self-reported normal	Self-reported normal	Self-reported normal	Self-reported normal	Self-reported normal	Self-reported normal

Tvol, testicular volume in mls; Ph, pubic hair; G, genital; B, breast stage; IU, international units; G2, genital stage 2; B2, breast stage 2; PHV, peak height velocity.
[a]Bone age estimated by the Greulich and Pyle method.
[b]Baseline values.
[c]Normal ranges for prepubertal boys given in parentheses. Induction of puberty where indicated was with intramuscular testosterone esters.

steroids and delayed bone age at presentation. In addition, these 10 probands displayed a typical growth pattern of self-limited DP, without compromised linear growth before puberty (Fig 2B). Although mean height SDS was below target height, the majority of patients fell within normal limits. At adult height, all but two probands (3.III.2 and 6.II.1) fell within normal limits for distance to target height (Table 4). Birth length, birth weight, timing of pubertal onset, and adult height of those with *IGSF10* mutations were similar to those of other affected DP individuals without *IGSF10* mutations from our cohort (Appendix Table S2).

In silico analysis of *IGSF10*

IGSF10 is a gene with thus far unclear function, mutations in which have not previously been associated with human disease. The IGSF10 protein has not previously been modeled by crystallography, so we therefore performed *in silico* analysis (Figs EV2–EV4 and EV5C, and Appendix Fig S2). *In silico* analysis of the protein reveals five defined regions (Fig 3A and B). Region I, in which the two N-terminal variants identified are located, contains leucine-rich repeats (LRR) 1–7 flanked by a LRR N-terminal (LRR Nt) and C-terminal (LRR Ct) cap. Region II is structurally disordered. Region III contains two immunoglobulin-like beta sandwich (Ig-like) domains. Region IV is structurally disordered. Region V, in which the two C-terminal variants are located, contains 10 further Ig-like domains (Fig EV5B and C). No clear evidence for a predicted transmembrane domain was provided by the *in silico* analysis. However, published data give evidence for a putative cleavage site with secretion of the N-terminal portion (Segev *et al*, 2004).

Table 4. Growth data of probands with *IGSF10* variants.

Case	Sex	Amino acid alteration	Height SDS at the age of 4 years	Height SDS at the age of 8/9 years	Target height	delta HSDS	Distance to target height at 4 years	Distance to target height at 8/9 years	Adult height SDS
1.II.1	M	p.Arg156Leu	−1.1	−1.3	−0.4	−0.2	0.7	0.9	−0.9
2.II.1	M	p.Arg156Leu	−0.2	−0.4	−0.7	−0.2	−0.5	−0.3	0.1
3.III.2	M	p.Glu161Lys	−0.5	−0.8	0.7	−0.3	1.2	1.5	−0.9
4.III.1	M	p.Glu161Lys	−0.4	−0.7	0.4	−0.3	0.8	1.1	0.0
5.II.1	M	p.Glu161Lys	−0.3	−0.2	0.9	0.1	1.2	1.1	0.9
6.II.1	M	p.Glu161Lys	−0.2	0	1.7	0.2	1.9	1.7	0.0
7.III.5	M	p.Asp2614Asn	−1	−1.5	−0.4	−0.5	0.6	1.1	−0.2
8.III.2	F	p.Asp2614Asn	−1.9	−2.2	−1	−0.3	0.9	1.2	−1.9
9.II.4	M	p.Asp2614Asn	−0.7	−1.6	−0.2	−0.9	0.5	1.4	−0.4
10.II.1	M	p.Glu2264Gly	−1.8	−2	−0.6	−0.2	1.2	1.4	−1.8

Height is expressed in SD score (SDS) for national reference data for Finland at 4 years of age, at either 8 years for girls or 9 years for boys, and at adult height. Normal limits: delta HSDS < 1.21, distance to target height at 4 years < 1.76, distance to target height at 8/9 years < 1.72, distance to target height at adult height < 1.44 (Saari *et al*, 2015).

IGSF10 N-terminal mutant proteins display pathogenic features with failure of extracellular secretion

The two N-terminal variants identified (p.Arg156Leu and p.Glu161Lys) are located in region I within LRR domains (Fig 2A). Both WT and mutant N-terminal protein fragments (668 aa in length) were expressed in HEK293 cells as demonstrated by Western blotting. However, while the GFP-tagged WT protein was detected in the conditioned media, neither mutant protein could be detected in their respective conditioned media. Moreover, a significant increase in mutant protein was detected in cell lysates, suggesting intracellular retention of these two mutants (Fig 3C).

Tissue expression studies localized *Igsf10* mRNA expression to the spatial and temporal window of GnRH neuronal migration

The expression of *Igsf10* mRNA in the nasal region of mouse embryos was analyzed from embryonic day (E) 10.5 to E17.5. During this developmental window, GnRH neurons emerge from the nasal placode (around E11) and then migrate toward the basal forebrain and hypothalamus (E12.5–E17.5) (Cariboni *et al*, 2007). *Igsf10* mRNA expression was undetectable in the nasal region or forebrain at E10.5 (Fig 4A). At E12.5, *Igsf10* was prominently expressed in the nasal mesenchyme (NM) with a decreasing gradient of expression from the area surrounding the vomeronasal organ (VNO) toward the nasal forebrain junction (NFJ), and absent in the VNO and olfactory epithelium (OE) (Fig 4B). At E12.5, GnRH neurons are exiting the VNO and migrating into an *Igsf10* strongly positive cell milieu in the NM (Fig 4B and C). At E14.5, GnRH cells and peripherin-positive olfactory axons are navigating among the *Igsf10*-positive NM cells (Fig 4D, G and H). *Cxcl12*, one of a plethora of molecules known to provide directional cues to GnRH neurons (Memi *et al*, 2013), is also expressed in the NM, although with an opposite gradient compared to *Igsf10* (Fig 4I). At E17.5, GnRH neurons are mainly located in the medial preoptic area (MPOA, Fig 4J). *Igsf10* signal was not detected in the hypothalamus at E17.5. *Igsf10*-positive cells were negative for isolectin b4 (marker of vasculature) at all developmental stages (Fig 4E, at E14.5). Incubation with the sense probe resulted in no signal at all stages. A similar expression pattern of *IGSF10* was detected in the nasal area of human embryos at 11 post-conceptual weeks (pcw) (Fig 4K–O), with GnRH neurons navigating among the *IGSF10*-positive NM cells.

Igsf10 knockdown *in vitro* leads to reduced migration of immature GnRH neurons

To investigate the functional role of *Igsf10* in the migration of GnRH neurons, we utilized a model of immortalized but migrating mouse GnRH cells (Radovick *et al*, 1991). These GN11 cells express neuronal markers and retain many features of immature GnRH-secreting neurons (Wetsel, 1995; Gore & Roberts, 1997), including a strong chemomigratory response *in vitro* (Maggi *et al*, 2000; Giacobini *et al*, 2002; Pimpinelli *et al*, 2003). We performed co-culture experiments of GN11 aggregates placed on confluent NIH3T3 monolayers. NIH3T3 cells, derived from a mouse embryonic fibroblast cell line, express high levels of endogenous *Igsf10*. The NIH3T3 cells were treated with scrambled- or Igsf10-shRNAs, the latter leading to highly reduced levels of *Igsf10* expression (Fig 5A). Migration of GN11 across Igsf10-shRNA-treated NIH3T3 cells was found to be significantly reduced compared to that across scrambled shRNA-treated NIH3T3 (two-tailed *t*-test, $n = 3$ for both Igsf10-shRNA and scr-shRNA with at least eight micromass replicates in each group, $P = 0.02422$) (Fig 5B and C, analysis in D).

Igsf10 knockdown *in vivo* results in perturbed migration and failed neurite extension of GnRH3 neurons

A transgenic zebrafish line, *Tg(gnrh3:EGFP)*, was used to visualize GnRH3 neurons and their projections. RT–PCR analysis indicated high and relatively equal expression of *Igsf10* mRNA at all timepoints tested during embryogenesis, from 48 h post-fertilization (hpf) onward (Fig 5E). Morpholino-modified antisense oligonucleotides (MOs) were used to repress *Igsf10* mRNA to assess the effect of *Igsf10* knockdown on the development of the GnRH3

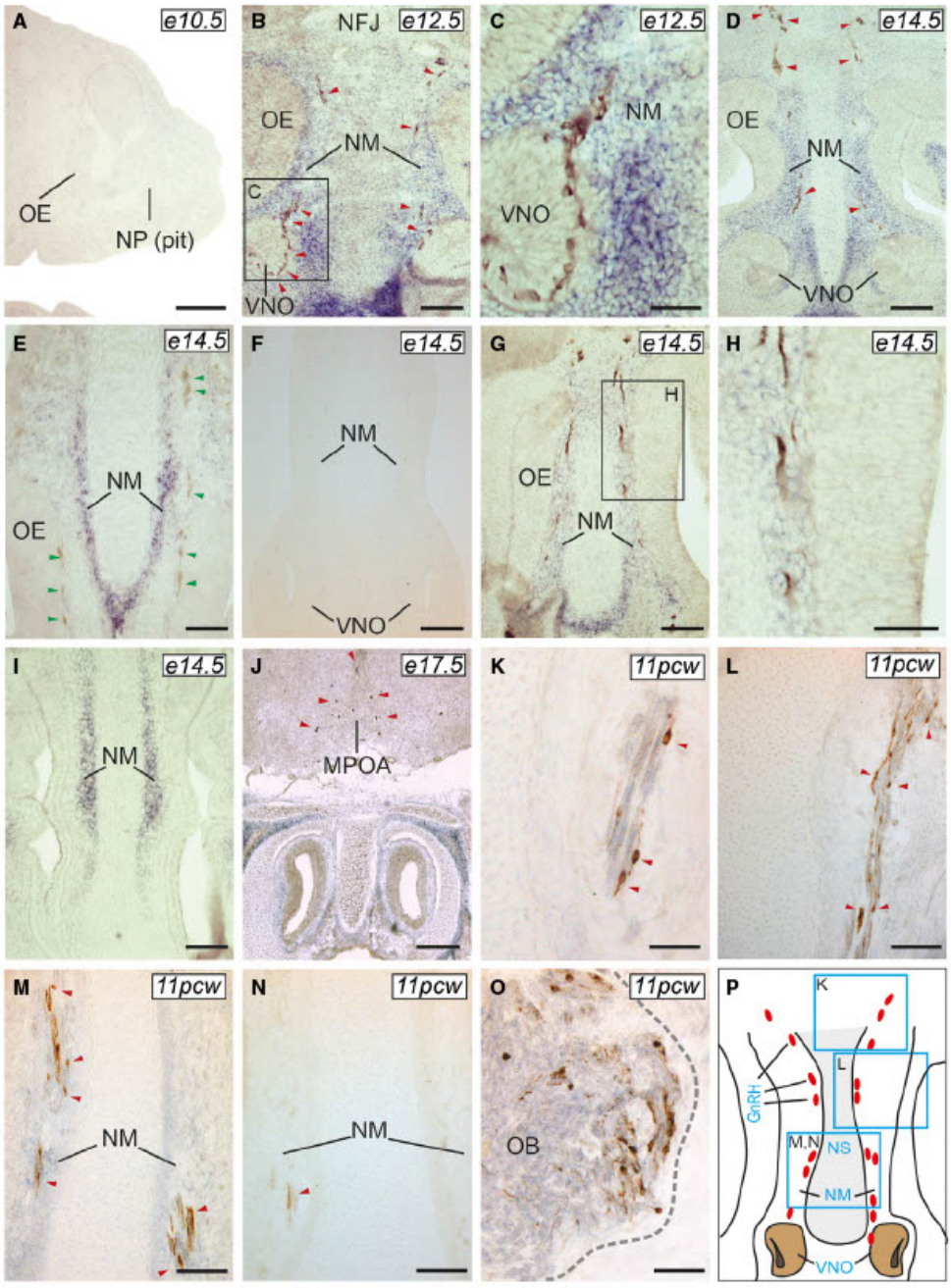


Figure 4.

Figure 4. Expression pattern of *Igsf10* mRNA in mouse and human developing brain.

A–J *Igsf10* expression was not observed at E10.5 (A, sagittal section), but readily detectable at E12.5, in the NM (B and C, frontal sections), and E14.5, along the migratory path of GnRH neurons (D, frontal section). *Igsf10*-positive cells were negative for isolectin b4 (a marker of vasculature) at all developmental stages (E, at E14.5). Incubation with the sense probe resulted in no signal at all stages (E14.5 shown in F). Peripherin-positive cells olfactory axons were extending within the *Igsf10*-positive NM (G and H). *Cxcl12* mRNA expression is shown in I (E14.5, frontal section). GnRH neurons are located in the MPOA by E17.5 (J, frontal section) in an *Igsf10*-negative area.

K–P In human 11pcw brains, *IGSF10* expression pattern was similar to that observed in mouse, with GnRH neurons interspersed in an *IGSF10*-positive NM (K–M, frontal sections, see also schematic in P). *IGSF10*-positive cells were also detected in the olfactory bulb (OB) (O). Sense probe resulted in no specific signal (N).

Data information: GnRH neurons are shown by red arrowheads and isolectin b4 by green arrowheads. NS: nasal septum. Scale bar, 250 μm (A), 100 μm (B, E, H), 50 μm (C, K–O), 200 μm (D, F, G, I, J).

system. Time-lapse analysis of MO-injected *Tg(gnrh3:EGFP)* embryos is shown in Fig 5F. At 48 hpf, GnRH3 neurons are normally seen as bilateral dots in the olfactory organ–olfactory bulb boundary. Over the following days, their projections extend caudally through the telencephalon to the hypothalamus. The strongest effect of morpholino injections was observed at 48 h. At this time-point, the percentage of embryos showing an abnormal GnRH3-neuron phenotype was higher in *Igsf10* splice-site-MO (Sp-MO)-injected embryos compared to relative controls, either injected with a mispair morpholino (control-MO) or uninjected (mean ± standard error of mean: Sp-MO 33% ± 3.6 vs. control-MO 11.7% ± 1.5 vs. uninjected 2.8 ± 1.3; one-way ANOVA, $n = 201/160/156$ for Sp-MO/Ctrl-MO/uninjected, $P = 0.000009$, Fig 5G). A similar effect was observed with the embryos injected with *Igsf10* ATG-MO. *Igsf10* knockdown affected both the guidance and the axonal outgrowth of GnRH3 neurons, which were unable to form compact clusters or extend projections to the hypothalamus.

Loss-of-function mutation in *IGSF10* in two patients with functional hypogonadotropic hypogonadism

To explore the possible role of mutations in *IGSF10* in conditions of GnRH deficiency, we carried out targeted exome sequencing of *IGSF10* in an adult cohort of 334 patients with HH due to Kallmann syndrome, idiopathic HH, or functional hypogonadism (hypothalamic amenorrhea (HA) or HA equivalent). This investigation showed that 10.2% of these patients carried a rare, predicted damaging variant in *IGSF10* (Appendix Table S3). Three loss-of-function variants (NM_178822: c.C352T: (rs142596318) p.R118*, NM_178822: c.G4804T: (rs79363433) p.E1602*, and NM_178822: c.7353_7354insATCA: (rs570110855) p.L2452 fs) were identified in a total of five patients, and 13 missense variants were identified in 29 patients from our HH cohort, all in the heterozygous state. In particular, all three of the loss-of-function variants were enriched in our HH cohort as compared to the ExAC database.

Two patients (out of 14 patients from this HH cohort with functional hypogonadism) were identified as heterozygous for a shared loss-of-function mutation in *IGSF10* (NM_178822: c.7353_7354insATCA: (rs570110855) p.L2452 fs). This variant is predicted to be deleterious with a high degree of confidence, with the expected loss of the last two Ig-like domains of the IGSF10 protein. It is a rare variant in the general population, with a minor allele frequency in the ExAC database of 0.01%. Both patients carrying the variant had adult-onset functional hypogonadism associated with environmental stressors. Patient 1 aged 31 years had a history of secondary amenorrhea induced by excessive exercise, and presented with failure to achieve spontaneous pregnancy. Patient 2 presented with adult-onset hypogonadism secondary to excessive weight loss and a subclinical eating disorder. Neither patient had a family history of HH, and both had normal pituitary imaging and normosmia. (A more detailed summary is given in Appendix Table S4.)

Discussion

The genetic control of the timing of puberty remains a fascinating and largely unsolved puzzle. The inheritance of DP is known to be under strong genetic influence with clear autosomal dominant segregation of the trait within families, and thus represents a useful basis for the investigation of puberty genetics. However, the genes responsible for DP have not been identified, other than in a small number of relatives of patients with HH (Lin *et al*, 2006; Pitteloud *et al*, 2006; Tornberg *et al*, 2011; Vaaralahti *et al*, 2011; Zhu *et al*, 2015). In this study, our findings indicate a role for *IGSF10* in the migration of GnRH neurons and highlight two pathogenic mutations in *IGSF10* as the causal factor for DP in six unrelated families. We have identified an additional two rare variants for unknown significance in four further families.

IGSF10, a gene of previously unclear function, is a member of the immunoglobulin superfamily. Loss-of-function mutations in another member of this superfamily, *IGSF1*, were recently identified in patients with X-linked central hypothyroidism (Sun *et al*, 2012). Notably, male patients with *IGSF1* mutations have a late increase in testosterone levels with a delayed pubertal growth spurt.

Our functional work specifies a likely role of *IGSF10* in the early migration of GnRH neurons. The development of the HPG axis is exceptional in that GnRH neurosecretory neurons develop in metazoan embryos outside of the central nervous system. Immature GnRH precursor neurons are first detectable in the olfactory placode in the nose from an early embryological stage (E11 in mice) and then begin a complex migration through the forebrain into the hypothalamus and preoptic areas (Cariboni *et al*, 2007; Wray, 2010).

An intact GnRH neurosecretory network is necessary for the correct temporal pacing of puberty, as demonstrated by animal models and the absence of pubertal development in patients with HH (Ojeda *et al*, 2006; Colledge *et al*, 2010; Palmert & Dunkel, 2012). Our tissue expression studies localized *IGSF10* mRNA expression to a spatial and temporal window when GnRH neurons are migrating through the nasal mesenchyme (E11.5–17.5 in mice) to the border with the telencephalon. Further evidence for the role of *IGSF10* in the regulation of GnRH neuronal migration was gained from our *in vitro* demonstration of reduced migration of immortalized GnRH neurons into a cell milieu with strongly reduced *Igsf10*

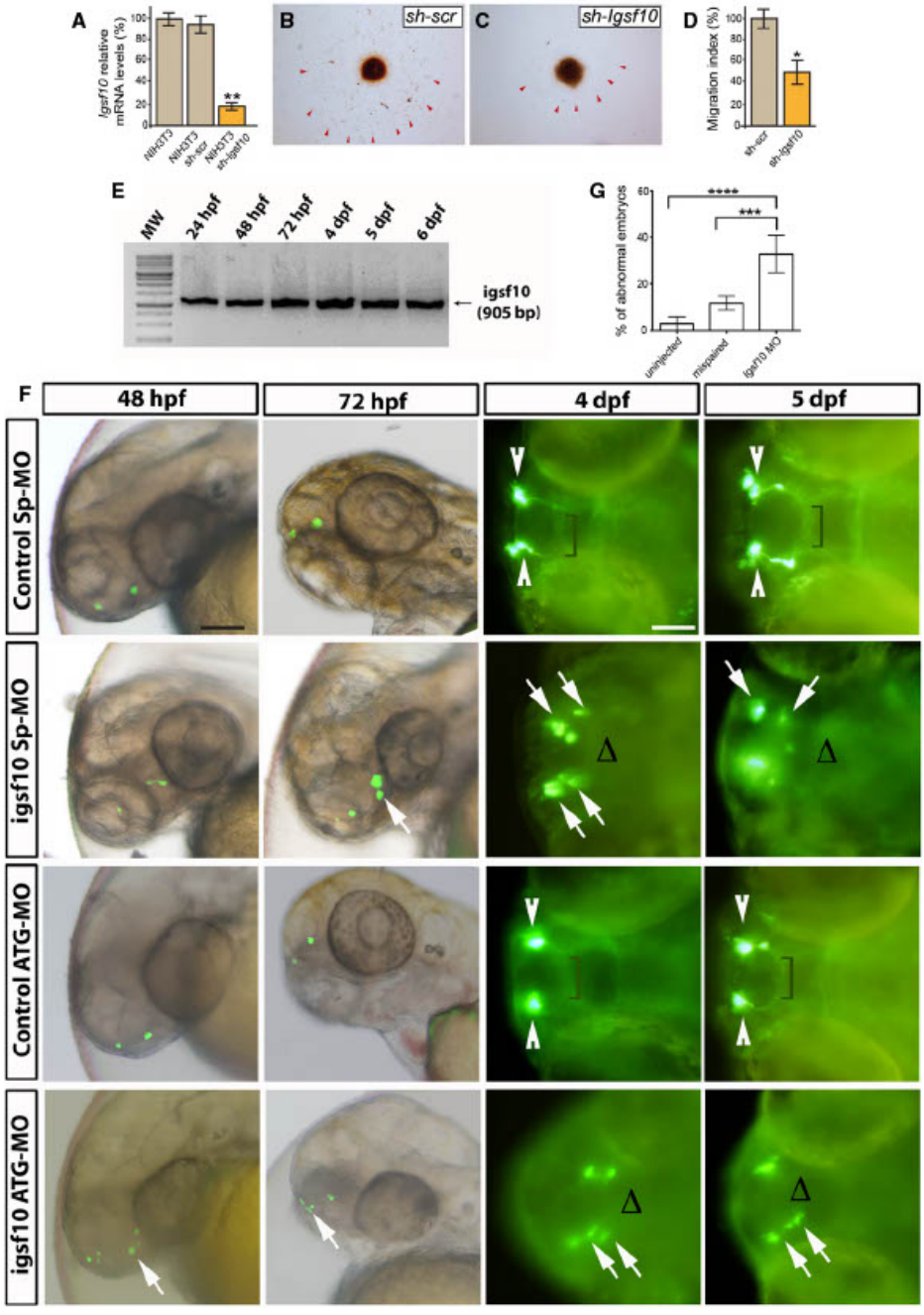


Figure 5.

Figure 5. Effect of *Igsf10* knockdown on GnRH neuronal migration.

A Levels of *Igsf10* expression in native NIH3T3 cells, and cells stably expressing a scrambled (sh-scr) or *Igsf10* (sh-Igsf10) shRNA.

B, C Migration of GN11 cells from aggregates into NIH3T3 sh-scr (B) or sh-Igsf10 (C).

D Analysis of the migration index (see Materials and Methods).

E RT–PCR analysis in zebrafish of total embryos and larvae showing the expression of *Igsf10* mRNA at different time-points. Expected band size: 905 bp.

F Representative time-lapse analysis of mispair control (control Sp-MO and control ATG-MO) and *Igsf10* morpholinos (Igsf10 Sp-MO and Igsf10 ATG-MO)-injected *Tg (gnrh3:EGFP)* embryos and larvae, at different time-points. Images show lateral view of live larvae head (48 and 72 hpf) and dorsal view (4 and 5 dpf) of live larvae head; anterior is left. White arrowheads indicate normal GnRH3 neuron clusters in the olfactory area of control embryos. Brackets show the extension of the projections toward the hypothalamus. White arrows indicate examples of abnormal GnRH3 neurons scattered in the olfactory area of morphants accompanied by the lack of projections toward the hypothalamus, indicated with Δ (hpf, hours post-fertilization; dpf, days post-fertilization). Scale bar, 100 μm (panels at 48 and 72hpf in F), 150 μm (panels at 4 and 5 dpf in F).

G Effect of *Igsf10* splice-site morpholino injection observed at 48 h. Quantification of the percentage of embryos showing an abnormal GnRH3-neuron phenotype was significantly higher in *Igsf10* splice-site-MO (Sp-MO)-injected embryos compared to relative controls.

Data information: (A, D) two-tailed *t*-test, $n = 3$ for each group $^{*}P = 0.02422$, $^{**}P = 0.00101$. (G) one-way ANOVA with Dunnett's *post hoc* test, $n = 201/160/156$ for Sp-MO/Ctrl-MO/uninjected $^{***}P = 0.00020$, $^{****}P = 0.00001$. (A, D, G) error bars represent mean ± SEM.

expression. These results were additionally validated by the use of a transgenic zebrafish model, where the depletion of *Igsf10* via injected MOs in zebrafish embryos resulted in perturbed migration and failed neurite extension of GnRH3 neurons toward the hypothalamus.

We hypothesize that immature GnRH neurons may respond to *IGSF10* signaling at the earliest stages of their migration, on exiting the VNO. Pathogenic *IGSF10* mutations, such as those identified here, will disrupt *IGSF10* signaling, potentially resulting in reduced numbers or mis-timed arrival of GnRH neurons at the hypothalamus, the latter leading to a functional defect in the formation of the GnRH neuroendocrine network. With this impaired GnRH system, there would follow an increased "threshold" for the onset of puberty, with an ensuing delay in pubertal timing. This relationship has also been demonstrated in adult Reeler mice, which have significantly fewer GnRH neurons in the hypothalamus and display a phenotype of delayed pubertal maturation (Cariboni *et al*, 2005).

GnRH neurons are known to have receptors for at least 20 neurotransmitters. Migratory GnRH neurons receive a plethora of guidance and movement-inducing messages during this journey, which are likely to be distinct depending on the stage of their migration (Tobet & Schwarting, 2006). Signals may act directly or indirectly through the extension of olfactory axons, as disruption of the nerve tract "scaffolds" themselves can disrupt GnRH migration (Gao *et al*, 2000). Gradients of chemokines may be particularly important for promoting the movement of GnRH neurons (Tobet & Schwarting, 2006). The pattern of *Igsf10* expression in the nasal mesenchyme shows a ventral to dorsal gradient between the VNO and olfactory bulbs, similar to a known axon-guidance gene, *Cxcl12*. *Cxcl12* and its receptor *Cxcr4* are necessary for the guidance of GnRH neurons toward the forebrain in both mouse (Toba *et al*, 2008) and zebrafish (Palevitch *et al*, 2010) models. In the absence of intact Cxcl12/Cxcr4 signaling, GnRH neurons fail to properly migrate out of the nasal mesenchyme during embryonic development.

The discovery of the role of *IGSF10* contributes an additional component to the highly complex system of secreted molecules and chemotactic gradients directing GnRH neuronal migration in the nasal region (Wray, 2010). The specific receptor(s) for secreted IGSF10 protein, on the GnRH neurons or elsewhere, remains to be determined. *IGSF10*, like *GnRH*, is also expressed in lung and other tissues, where its function is as yet unknown. A degree of redundancy with or compensation by other chemokines, such as has been postulated for the GnRH neuron migratory cue semaphorin-4D, as well as the action of potential "protective factors", is feasible (Wray, 2010).

Furthermore, mutations in *IGSF10* may contribute to the phenotype of other forms of secondary hypogonadism. Mutations that perturb GnRH neuronal migration, including in *KAL1* and *PROKR2*, are already known to cause both HH and HA (Caronia *et al*, 2011). Moreover, the same authors found that 25% of patients with HA in their study had DP. DP is commonly found in relatives of patients with HH and Kallmann syndrome. These findings imply that disruption to the GnRH network may result in a spectrum of phenotypes from DP through HA to complete GnRH deficiency. Specifically, this clinical variability can result from mutations in genes such as *KAL1*, *PROKR2*, and now *IGSF10*, which may lead to late arrival or reduced numbers of GnRH neurons to the hypothalamus, thus compromising the function of the GnRH network. Moreover, an increasing burden of mutations may produce a more severe outcome, with perhaps one mutation leading to DP or HA, while two or even more mutations may be required to lead to a phenotype of HH or KS. Our finding of loss-of-function mutations in *IGSF10* in two out of 14 individuals sequenced with HA and HA equivalent adds weight to this suggestion, although sequencing of *IGSF10* in a larger cohort of HA patients would help to confirm this assertion. However, although rare, predicted pathogenic variants in *IGSF10* were enriched in our HH cohort as compared to the general population, the presence of such variants in the public databases suggests that these variants alone are not sufficient to cause complete idiopathic HH or KS, again in keeping with the above hypothesis.

In conclusion, our findings strongly support the contention that mutations in *IGSF10* cause delayed puberty in humans, through dysregulation of GnRH neuronal migration during embryonic development. Moreover, such mutations may also underlie susceptibility to the phenotype of hypothalamic amenorrhea. Overall, this represents a new causal mechanism for self-limited DP and reveals a shared pathophysiology between DP and other forms of functional hypogonadism.

Materials and Methods

Patients

Initial whole exome sequencing was performed in 18 probands with self-limited DP and their affected and unaffected family members

from our previously described cohort (Wehkalampi *et al*, 2008a). In brief, patients referred with DP to specialist pediatric care in central and southern Finland between 1982 and 2004 were identified. All patients ($n = 492$) met the diagnostic criteria for self-limited DP, defined as the onset of Tanner genital stage II (testicular volume > 3 ml) >13.5 years in boys or Tanner breast stage II > 13.0 years in girls (i.e., two SD later than average pubertal development) (Palmert & Dunkel, 2012). Medical history, clinical examination, and routine laboratory tests were reviewed to exclude those with chronic illness. HH, if suspected, was excluded by spontaneous pubertal development at the follow-up. In the 50% of patients from the cohort who choose to have pubertal induction via the use of exogenous sex steroids, all patients were followed up once off treatment until the point of full pubertal development (Tanner stage G4+ or B4+) to ensure that pubertal development did not arrest off treatment.

Families of the DP patients were invited to participate, with information about medical history and pubertal timing obtained by structured interviews and from archived height measurement records. The criteria for the diagnosis of DP in family members were one of the following three: (i) age at takeoff, or (ii) peak height velocity (phv) occurring 1.5 SD beyond the mean, that is, age at takeoff exceeding 12.9 and 11.3 years, or age at phv exceeding 14.8 and 12.8 years in males and females, or (iii) age at attaining adult height more than 18 or 16 years, in males and females, respectively (Wehkalampi *et al*, 2008a). All family members were assigned a clinical status of affected, unaffected, or unknown. Those with unknown status were either too young to diagnose or had insufficient growth data available.

Written informed consent was obtained from all participants.

A hypogonadotropic hypogonadism (HH) cohort of patients with idiopathic hypogonadotropic hypogonadism (IHH, $n = 158$), Kallmann syndrome (KS, $n = 162$), and hypothalamic amenorrhea (HA) or HA equivalent ($n = 14$) were collected through three coordinating European centers (Newcastle upon Tyne Hospital, Newcastle, UK; Bicêtre University Hospital, Paris, France; and Robert Debré Hospital, Paris, France). Patients with IHH all had absent or incomplete puberty at 18 years of age, low or normal serum gonadotropin levels, low serum estradiol or testosterone levels, otherwise normal anterior pituitary function, and normal results on neuroimaging. KS patients had absent or incomplete puberty at 18 years of age, low or normal serum gonadotropin levels, low serum estradiol or testosterone levels, otherwise normal anterior pituitary function, and anosmia diagnosed clinically and by radiological evidence of olfactory bulb disruption. Patients with HA or HA equivalent had a history of spontaneous pubertal development with secondary amenorrhea in women for 6 months or more, low or normal gonadotropin levels, low serum estradiol or testosterone levels, and one or more predisposing factors. These factors included excessive exercise (> 5 h per week), or other stress, loss of more than 15% of body weight, and either evidence of a subclinical eating disorder, or another cause of dietary restriction (eg. self-perceived "intolerance" to certain foods). Patients with frank anorexia nervosa or BMI < 17 kg/m^2 were excluded.

DNA sequencing and bioinformatics

Whole exome sequencing was initially performed on DNA extracted from the peripheral blood leukocytes of 111 individuals from the 18 most extensive families from our cohort (a total of 76 individuals with DP: male = 53 and female = 23, and 35 controls, male $n = 13$ and female $n = 22$), with exome capture on a Nimblegen V2 platform or Agilent V5 platform and sequencing on the Illumina Hiseq 2000. The exome sequences were aligned to the UCSC hg19 reference genome. Picard tools and the genome analysis toolkit were used to mark PCR duplicates, realign around indels, recalibrate quality scores, and call variants.

Variants were analyzed further and filtered for potential causal variants using filters for quality control, predicted functional annotation, minor allele frequency, segregation with trait, variants in multiple families, and biological relevance (Fig 1). Quality control included thresholds for call quality, read depth, and upstream pipeline filtering. Predicted functional annotation involved prioritizing nonsense, exonic missense, splice-site variants, structural or promoter changes, or variants deleterious to a microRNA. Filtering by MAF entailed including those variants with minor allele frequency (MAF) < 2.5% in the 1000 Genomes database, the NHLBI exome variant server, Sequencing Initiative Suomi (sisu.fimm.fi), and dbSNP databases. Segregation with trait refers to variants present in $\geq n-1$ affected individuals (where n = the number of affected individuals in a given pedigree) and not present in more than one unaffected individual being retained. Multiple family filtering involved retaining those variants seen in more than one family from the cohort sequenced, or with different variants in the same gene in more than one family. Biological relevance filtering allowed prioritization of those remaining variants under three criteria: (i) Variants in genes known to be relevant to the phenotype of HH; (ii) variants in genes in linkage disequilibrium (r^2 no limit, D prime > 0.8) with loci associated with genome-wide association studies of age at menarche (Elks *et al*, 2010); and (iii) variants in genes with potential biological significance, using the tools Ingenuity Variant Analysis (QIAGEN Redwood City, www.qiagen.com/ingenuity), Genego MetaCore (Thomson Reuters), Ensembl variant effect predictor, and Annovar (Wang *et al*, 2010).

Targeted exome sequencing using a Fluidigm array for library preparation followed by Illumina Miseq sequencing of 28 candidate genes identified post-filtering was then performed on DNA extracted from the peripheral blood leukocytes of a further 42 families from the same cohort (288 individuals, 178 with DP and 110 controls, Fig 1). Variants post-targeted resequencing were filtered using the same criteria as the whole exome sequencing data: on the basis of quality control, predicted functional annotation, minor allele frequency, and segregation with trait.

Rare variant burden testing was performed for all 28 genes post-targeted resequencing. Fisher's exact test was used to compare the prevalence of deleterious variants in our cohort with the Finnish population, using the ExAC Browser (Exome Aggregation Consortium (ExAC), Cambridge MA: http://exac.broadinstitute.org, accessed September 2015). For each of the 28 genes, all variants from the ExAC database with minor allele frequency < 2.5%, predicted to be deleterious by both Polyphen-2 (Adzhubei *et al*, 2010) and SIFT (Kumar *et al*, 2009), were included in the analysis, with each family in our cohort represented by the proband only. A multiple comparison adjustment was applied to the set of 28 *P*-values *post hoc* using the Benjamini and Hochberg method (Benjamini *et al*, 2001). Variants identified following filtering

pipelines were confirmed by conventional Sanger sequencing and screened in a further 100 controls from our cohort via Sanger sequencing.

The HH cohort ($n = 334$) was screened via targeted exome sequencing for mutations in *IGSF10* using a Fluidigm array for library preparation followed by Illumina Miseq sequencing, with the same filtering pipeline used for the whole and previous targeted exome sequencing data. Variants identified were confirmed by conventional Sanger sequencing.

Growth pattern analysis

The pattern of prepubertal growth in the 10 probands was analyzed using two parameters: (i) HSDS distance from target height (TH) at the ages of 4 and 8 (girls) or 9 (boys) years (TH formula = 0.791 × mean parental height SDS −0.147 for girls and 0.886 × mean parental height SDS −0.071 for boys) and (ii) change in height SDS (ΔHSDS) between the ages of 4 and 8/9 years. Normal values for the two parameters, based on data from > 70,000 healthy Finnish children, have been previously published (Saari *et al*, 2015).

In silico analysis

The amino acid sequence for human *IGSF10* was retrieved from UNIPROT database (id Q6WRI0). Homology modeling was used to determine the 3D structure using two high-performing protein structure prediction servers: Phyre2 (Kelley & Sternberg, 2009; Jefferys *et al*, 2010) and I-Tasser (Roy *et al*, 2010). The signal peptide cleavage site was calculated using SignalP4.1 (Petersen *et al*, 2011). The following interactions involved in protein stability were considered: (i) salt bridges, defined as at least one pair of atoms on oppositely charged groups within a 4.5Å distance; (ii) hydrogen bonds (H-bond), defined as a donor–acceptor distance ≤ 2.5 Å and an angle at the acceptor ≥ 90°; and (iii) disulfide bridge (S-S bridge) defined as the side chains of two cysteines at a 3.0 Å distance. Pairs of cysteines at a greater distance were also considered as potentially forming an S-S bridge when found to be reasonably close (Cα-Cα distance < 10 Å). We used Cα distance to allow for errors in side-chain placement and a relatively high threshold to accommodate possible deviation of the backbone from native. The protein electrostatic potential was calculated using PBEQ program (Jo *et al*, 2008), which computes the protein electrostatic potential by solving the Poisson–Boltzmann equation. Protein structures were visualized using the Pymol visualization program (http://www.pymol.org/).

Constructs and protein expression

An N-terminal fragment encoding the first 668 aa of *IGSF10* gene (RefSeq NM_178822.3) was cloned into a pcDNA-EGFP (Addgene plasmid #13031), and p.Arg156Leu and p.Glu161Lys variants were generated using PCR-mediated mutagenesis (Quickchange II, Agilent Technologies) according to the manufacturer's instructions and verified by sequencing.

HEK293 (sourced from ATCC) were transfected with wild-type or mutant plasmids using polyethylenimine (Sigma) via a standard protocol. Cells were checked for mycoplasma contamination (MycoAlert Detection Kit, Lonza) on a monthly basis and were contamination-free. Transfected cells were selected with G418 (1 mg/ml, Sigma). Once stable cell lines were established, the cells were cultured for 24 h in 6-well plates to full confluency in 1 ml of media. This process was repeated on three separate occasions with cells from the same stable cell lines, to produce technical replicates. Conditioned media from each well were removed for analysis before cell lysis with RIPA buffer (Sigma) supplemented with protease inhibitors (Complete Mini, Roche). Lysates and conditioned media were cleared by centrifugation at 16,060 *g* at 4°C for 10 min. Equal amounts of lysates and conditioned media were size-separated (NuPage BisTris gels 4–12%) and transferred to nitrocellulose membranes (Promega). Post-transfer, the blots were stained with Ponceau red to assess equal loading, blocked with PBS containing 0.1% Tween-20 and 5% nonfat dry milk, and incubated in a 1:500 dilution of mouse monoclonal anti-GFP (Roche, 1DB-001-0000570956) antibody or a 1:3000 dilution of mouse monoclonal anti-glyceraldehyde-3-phosphate dehydrogenase (GAPDH, Santa Cruz Biotechnology, 1DB-001-0000183498) for 12 h at 4°C. After washes, the membranes were incubated with goat anti-mouse IRDye680 (1:10000 dilution; Licor). Immunoblots were scanned with, and protein relative amounts were calculated by, the Odyssey ® Fc Imaging System (Licor).

Nonradioactive *in situ* hybridization (NR-ISH)/Immunohistochemistry

E10.5 to E17.5 mouse embryos were collected from timed crosses of C57BL/6 mice. The morning of the vaginal plug was designated 0.5 days; 11 post-conceptual weeks (pcw) human brains were obtained from the MRC-Wellcome Human Development Biology Resources (HDBR—Institute of Genetics Medicine, Newcastle, UK). Tissues were fixed in 4% paraformaldehyde (PFA) in PBS, cryoprotected in 30% sucrose, and frozen in OCT compound (VWR); 12-μm-thick serial sagittal and coronal sections were collected on Superfrost Plus slides (VWR).

Mouse and human *Igsf10* were PCR-amplified from brain cDNAs using the following primers: m*Igsf10* FOR: 5′-GCAAGAAGGAAAGAATCCCC -3′, REV: 5′- GATTCGCCCATCCTCACTAA -3′; h*IGSF10* FOR: 5′-TCAGGAGCTTGACACGATTG-3′, REV: 5′- CTGCGGTGTTTCACTAAGCA-3′. Amplified cDNAs were cloned into the dual promoter vector pGEM-T easy (Promega) and linearized with the appropriate restriction enzymes. Mouse Cxcl12 probe was from Memi *et al* (2013). Probe preparation and *in situ* protocol were performed as previously in Guasti *et al* (2011).

When co-labeling was desired, after *in situ*, the sections were incubated with primary antibodies (anti-GnRH, Immunostar; anti-peripherin, Merck-Millipore; anti-isolectin B4, Sigma) diluted 1:1000 in PBS–Triton 0.1%, overnight at room temperature (RT) as used in Cariboni *et al* (2011). After three washes with PBS–Triton 0.1%, the slides were incubated for 2 h at RT with biotin-conjugated goat secondary antibodies (Vector Laboratories), diluted 1:300 in PBS and, after further washes, with the avidin–biotin complex (ABC staining kit, Vector Laboratories). The sections were reacted with 4′,6′-diamino-2-phenylindole (DAPI, Vector Laboratories) and mounted as above.

Images were acquired using a Leica DM5500B microscope (Leica, Nussloch, Germany), equipped with a DCF295 camera (Leica) and DCViewer software (Leica), and then processed with Abode Photoshop CS6 and Adobe Illustrator CS6.

Migration experiments

Igsf10 silencing was achieved in NIH3T3 cells (sourced from ATCC) using SureSilencing shRNA plasmids (QIAGEN) according to the manufacturer's instructions. Stable lines were obtained by puromycin treatment (Life Technologies, 2 g/ml). Cells were checked for mycoplasma contamination (MycoAlert Detection Kit, Lonza) on a monthly basis and were contamination-free. The level of *Igsf10* knockdown was assessed by RT–qPCR. RNA extraction (RNeasy Mini kit, QIAGEN) and cDNA synthesis (M-MLV Reverse Transcriptase, Promega) were performed according to the manufacturers' instructions. RT–qPCR was performed in a 10-µl reaction mixture containing 2 µl cDNA template, 5 µl 2 × $SYBR_{GREEN}$ I Master Mix (KAPA Biosystems), 0.2 µl low ROX (KAPA Biosystems), 0.5 µl primers (10 µM forward + reverse), and 2.3 µl nuclease-free H_2O. *Gapdh* was used as the endogenous housekeeping gene. The real-time PCR was performed using an Mx3000 Thermocycler (Stratagene) using the following primers and conditions: m*Igsf10*: FOR, 5′-CTGGGGAGTCCAATTGCTGT-3′ and REV, 5′-GCTGCCTTTGCTG ACATC-3′ (18 bp); *Gapdh*: FOR, 5′-TGCACCACCAACTGCTTAG-3′ and REV, 5′-GGATGCAGGGATGATGTTC-3′. Quantitative RT–qPCR was set up in triplicate. After an initial denaturation step of 3 min at 95°C, PCR cycling was performed for 40 cycles of 95°C for 3 s, 55°C for 20 s and 72°C for 1 s, followed by 1 cycle of 1 min at 95°C, 55°C for 30 s and 95°C for 30 s. Silencing was achieved at 80% in *Igsf10*-silenced NIH3T3 cells compared to cells expressing the scrambled shRNA.

Micromass cultures were obtained with the immortalized GnRH-expressing cell line GN11. Confluent monolayer cultures of GN11 were released by trypsin–EDTA and resuspended in growth medium. Micromasses were obtained by pipetting 20 µl (4.0×10^5 cells) of cell suspension onto the lid of a 10-cm petri dish and incubating for 48 h in a 5% CO_2 incubator at 37°C. Micromasses were then placed onto confluent scrambled and *Igsf10*-silenced NIH3T3 cells for 7 days. The cells were then fixed in 4% PFA, and GN11 migration into NIH3T3 was assessed by staining with anti-GnRH antibodies, as used in Cariboni *et al* (2011). Migration index was calculated by assessing GnRH-stained cells using the ImageJ software (NIH). Statistical significance was evaluated by ANOVA and Student's *t*-test. $^*P < 0.05$ and $^{**}P < 0.01$ were taken to be significant. Data are presented as mean ± standard error of mean and expressed as percentage reduction of migration, taking untreated or scrambled controls as 100%. Each experiment was performed using at least 12 micromasses/group and repeated three times.

Zebrafish investigations

Animals

A transgenic line, *Tg(gnrh3:EGFP)*, was used to visualize GnRH3 neurons and their projections (Abraham *et al*, 2008). Wild-type and transgenic zebrafish embryos were generated by natural mating. Both AB and TL strains were used for the experiments. Embryos were raised in fish water with methylene blue (0.3 p.p.m.) in petri dishes at 28°C in a 12-h light/12-h dark cycle. Developmental stages are expressed in hours post-fertilization (hpf) or days post-fertilization (dpf). All embryos used in experiments were 5 dpf or younger. Pigmentation was prevented by adding 0.2 mM phenylthiourea (PTU) to the water at 24 hpf.

RT–PCR analysis

Total RNA was isolated from embryos and larvae at different stages of embryogenesis using EZ RNA Total RNA Isolation kit (Biological Industries, Beit Haemek, Israel). First-strand cDNA was synthesized from 1 µg of total RNA by qScript cDNA kit (Quanta Biosciences, Gaithersburg, MD, USA) according to the manufacturer's protocol. Primers for Igsf10: FOR, 5′-TTGGCTACAGTCCCGATTTC-3′ and REV, 5′-AAATTTTGCTGGGACGAATG-3′.

Morpholino-knockdown experiments

Morpholino-modified antisense oligonucleotides (MOs) (Gene Tools, Philomath, OR, USA) were used to repress *Igsf10* mRNA to assess the effect on the development of the GnRH3 system. A splice-site-MO (Igsf10 Sp-MO 5′-GCCTGTTGTAGGTTTTACCCCAGGT-3′, 25 bp) and an ATG-MO (Igsf10 ATG-MO 5′-GGGAATCCGCTGCTGGGTCA-CACAT-3′, 25 bp) were designed. Quantities of 1.5pmol/embryo (Igsf10 Sp-MO) or 1pmol/embryo (*Igsf10* ATG-MO) were injected into *Tg(gnrh3:EGFP)* embryos immediately after fertilization. In all experiments, Igsf10 Sp-MO- or ATG-MO-injected embryos were compared with embryos injected with the same amount of mispair control (control Sp-MO 5′-GCGTCTTGTACGTTTTACCCGACGT-3′ or control ATG-MO 5′-GAGAATACGCTACTGGGTAACAAAT-3′, both 25 bp) at the same developmental stage.

Two sets of experiments (biological replicates) were performed and each experiment consisted of three groups: Sp-MO injected, Ctrl-MO injected, and uninjected. No statistical method was used to predetermine sample size. Rather, sample size was based on preliminary data and observed effect sizes. Dead embryos or embryos that were GFP negative at 48 hpf were excluded from the analysis. Considering that MO injection can cause a death rate of 10–25% in 24-h-old embryos, we injected a large number of embryos in order to establish a statistical significance. In the first experiment, the mortality rate in the injected embryos, Sp-MO and Ctrl-MO, was 11% and in the second, it was 25%. The mortality rate in the uninjected embryos was 2.5 and 5.7%, respectively.

In the first experiment, 353 embryos were injected with Sp-MO, 260 embryos were injected with Ctrl-MO, and 244 embryos left uninjected: 146 of 353, 108 of 260, and 100 of 244 embryos were available and GFP positive for analysis at 48 hpf. In the second experiment, 186 embryos were injected with Sp-MO, 157 embryos were injected with Ctrl-MO, and 122 embryos left uninjected: 55 of 186, 52 of 157, and 56 of 122 embryos were available and GFP positive for analysis at 48 hpf. Sex of embryos was undetermined (separate sexes can be detected only after 21–23 dpf).

Injected larvae were checked daily and the EGFP expression was monitored under SZX12 fluorescent dissecting microscope (Olympus, Tokyo, Japan) for 5 days. Quantification of the percentage of embryos showing an abnormal GnRH3-neuron phenotype in *Igsf10* splice-site-MO (Sp-MO)-injected embryos compared to relative controls, injected either with a mispair morpholino (control-MO) or uninjected, was assessed by mean ± standard error of mean, tested by one-way ANOVA followed by Dunnett's *post hoc* test.

Efficacy of Igsf10 Sp-MO was evaluated by RT–PCR. Injected and uninjected embryos were sampled at 48 hpf and total RNA was extracted (EZ RNA Total RNA Isolation kit, Biological Industries, Beit Haemek, Israel); 1 µg of RNA was retro-transcribed using qScript cDNA kit (Quanta Biosciences, Gaithersburg, MD, USA). Specific primers corresponding to zebrafish *Igsf10* exons 2 and 3

(Igsf10-E2FOR 5′-GCGGATTCCCGTCTGCTATGG-3′ and Igsf10-E3REV 5′-TGCAGAGATGTGAGGCCACTGAAC-3′) were used for PCR amplification. PCR products were separated on an agarose gel, extracted from gel using HiYield Gel/PCR DNA Fragments Extraction Kit (RBC Bioscience, New Taipei City, Taiwan), and sequenced. Sequencing analysis demonstrated the insertion of 97 nucleotides within the intron 2 of Igsf10 Sp-MO-injected embryos, which would result in a frameshift and produce a dysfunctional Igsf10 protein in these embryos (Appendix Fig S3).

Study approval

The study protocol was approved by the Ethics Committee for Pediatrics, Adolescent Medicine and Psychiatry, Hospital District of Helsinki and Uusimaa (and extended to encompass Kuopio, Tampere, and Turku University Hospitals). UK ethical approval was granted by the London-Chelsea NRES Committee. The study was conducted in accordance with the guidelines of the Declaration of Helsinki. All patients gave informed written consent prior to inclusion in the study. All mouse experiments were conducted under the Animal (Scientific Procedures) Act 1986, Project Licence PPL 70/8269. The human embryonic and fetal material was provided by the Joint MRC/Wellcome Trust Grant# 099175/Z/12/Z Human Developmental Biology Resource (http://hdbr.org).

Statistics

For rare variant burden testing, Fisher's exact test was used with a multiple comparison adjustment applied *post hoc* using the Benjamini and Hochberg method (Benjamini *et al*, 2001). A threshold of $P < 0.025$ was taken as significant. Fisher's exact test was also used for the analysis of HH cohort allele frequencies. The unpaired *t*-test (two-tailed) was used for statistical analysis of auxological data (Appendix Table S2).

For all experiments, data are expressed as the mean ± SEM. To determine the statistical significance, we used the unpaired *t*-test (two-tailed) or, for multiple comparisons, a one-way ANOVA followed by a Dunnett's *post hoc* test. A *P*-value of less than 0.05 was considered statistically significant. Statistical analysis was performed using GraphPad Prism4 (GraphPad Software).

Expanded View for this article is available online.

Acknowledgements
SRH is funded by The Wellcome Trust (102745), Rosetrees Trust (M222), and the Barts and the London Charity (417/1551). LG and GRB are funded by the Biotechnology and Biological Sciences Research Council (BB/L002671/1). LD is partly supported by the Academy of Finland (14135). MRB, HRW, and CPC are funded by the National Institutes for Health Research (NIHR), and this work forms part of the portfolio of translational research of the NIHR Biomedical Research Unit at Barts. AD is funded by the MRC (MR/K021613/1). AC is funded by the Telethon Foundation (GP13142); VA is partly supported by a COST STSM (BM1105-16145) and a Travel Grant sponsored by Developmental Journal (The Company of Biologists Limited). The human embryonic and fetal material was provided by the Joint MRC/Wellcome Trust Grant# 099175/Z/12/Z Human Developmental Biology Resource (http://hdbr.org). We would especially like to thank all patients and their families who participated in the study.

The paper explained

Problem
Early or late pubertal onset affects up to 5% of adolescents and is associated with adverse health and psychosocial outcomes. Self-limited delayed puberty segregates in families, suggesting strong genetic influences, but the underlying mechanistic and genetic determinants of the pathogenesis of delayed puberty are not understood.

Results
Using whole and candidate exome sequencing in a large cohort with familial delayed puberty, we have identified a novel gene, *IGSF10*, mutations in which lead to self-limited delayed puberty. Our functional annotation of *IGSF10* suggests that it is a component of the complex system of migratory cues, which together guide GnRH neurons from their origin in the nasal placode toward the hypothalamus during embryogenesis. Loss-of-function mutations in *IGSF10* were also identified in patients with hypothalamic amenorrhea.

Impact
Here, we identify a new causal mechanism for self-limited delayed puberty, through defects in embryological migration of GnRH neurons to the hypothalamus, and reveal a shared pathophysiology between simple delayed puberty and other forms of functional hypogonadism.

Author contributions
SH, LG, and LD planned the genetic and experimental work. SH, LG, AM, and GR-B carried out the *in vitro* analysis. VA and YG carried out the zebrafish work. CC, HW, and MB gave bioinformatics and statistical support. AD and MS carried out *in silico* analyses. LD, KW, RQ, NdR, JY, and AG-M provided clinical samples and data. SH, LG, LD, AC, MB, HS, and LM contributed to data review and interpretation.

Conflict of interest
All authors declare that they have no conflict of interest.

For more information
IGSF10
http://www.uniprot.org/uniprot/Q6WRI0
http://www.genecards.org/cgi-bin/carddisp.pl?gene=IGSF10
Delayed puberty
http://patient.info/doctor/delayed-puberty

References

Abraham E, Palevitch O, Ijiri S, Du SJ, Gothilf Y, Zohar Y (2008) Early development of forebrain gonadotrophin-releasing hormone (GnRH) neurones and the role of GnRH as an autocrine migration factor. *J Neuroendocrinol* 20: 394 – 405

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR (2010) A method and server for predicting damaging missense mutations. *Nat Methods* 7: 248 – 249

Beate K, Joseph N, de Nicolas R, Wolfram K (2012) Genetics of isolated hypogonadotropic hypogonadism: role of GnRH receptor and other genes. *Int J Endocrinol* 2012: 147893

Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I (2001) Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 125: 279 – 284

Bianco SD, Kaiser UB (2009) The genetic and molecular basis of idiopathic hypogonadotropic hypogonadism. *Nature Rev Endocrinol* 5: 569 – 576

Cariboni A, Rakic S, Liapi A, Maggi R, Goffinet A, Parnavelas JG (2005) Reelin provides an inhibitory signal in the migration of gonadotropin-releasing hormone neurons. *Development* 132: 4709 – 4718

Cariboni A, Maggi R, Parnavelas JG (2007) From nose to fertility: the long migratory journey of gonadotropin-releasing hormone neurons. *Trends Neurosci* 30: 638 – 644

Cariboni A, Davidson K, Dozio E, Memi F, Schwarz Q, Stossi F, Parnavelas JG, Ruhrberg C (2011) VEGF signalling controls GnRH neuron survival via NRP1 independently of KDR and blood vessels. *Development* 138: 3723 – 3733

Caronia LMMC, Welt CK, Sykiotis GP, Quinton R, Thambundit A, Avbelj M, Dhruvakumar S, Plummer L, Hughes VA, Seminara SB *et al* (2011) A genetic basis for functional hypothalamic amenorrhea. *N Engl J Med* 364: 215 – 225

Chun S, Fay JC (2009) Identification of deleterious mutations within three human genomes. *Genome Res* 19: 1553 – 1561

Colledge WH, Mei H, d'Anglemont de Tassigny, X (2010) Mouse models to study the central regulation of puberty. *Mol Cell Endocrinol* 324: 12 – 20

Cooper GM, Stone EA, Asimenos G, Program NCS, Green ED, Batzoglou S, Sidow A (2005) Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res* 15: 901 – 913

Day FR, Elks CE, Murray A, Ong KK, Perry JR (2015) Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank study. *Sci Rep* 5: 11208

DiVall SA, Radovick S (2008) Pubertal development and menarche. *Ann N Y Acad Sci* 1135: 19 – 28

Elks CE, Perry JR, Sulem P, Chasman DI, Franceschini N, He C, Lunetta KL, Visser JA, Byrne EM, Cousminer DL *et al* (2010) Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nat Genet* 42: 1077 – 1085

Gajdos ZK, Hirschhorn JN, Palmert MR (2009) What controls the timing of puberty? An update on progress from genetic investigation. *Curr Opin Endocrinol Diabetes Obes* 16: 16 – 24

Gao C, Noden DM, Norgren RB Jr (2000) LHRH neuronal migration: heterotypic transplantation analysis of guidance cues. *J Neurobiol* 42: 95 – 103

Giacobini P, Giampietro C, Fioretto M, Maggi R, Cariboni A, Perroteau I, Fasolo A (2002) Hepatocyte growth factor/scatter factor facilitates migration of GN-11 immortalized LHRH neurons. *Endocrinology* 143: 3306 – 3315

Gore AC, Roberts JL (1997) Regulation of gonadotropin-releasing hormone gene expression in vivo and in vitro. *Front Neuroendocrinol* 18: 209 – 245

Guasti L, Paul A, Laufer E, King P (2011) Localization of Sonic hedgehog secreting and receiving cells in the developing and adult rat adrenal cortex. *Mol Cell Endocrinol* 336: 117 – 122

He C, Zhang C, Hunter DJ, Hankinson SE, Buck Louis GM, Hediger ML, Hu FB (2010) Age at menarche and risk of type 2 diabetes: results from 2 large prospective cohort studies. *Am J Epidemiol* 171: 334 – 344

Jefferys BR, Kelley LA, Sternberg MJ (2010) Protein folding requires crowd control in a simulated cell. *J Mol Biol* 397: 1329 – 1338

Jo S, Vargyas M, Vasko-Szedlar J, Roux B, Im W (2008) PBEQ-Solver for online visualization of electrostatic potential of biomolecules. *Nucleic Acids Res* 36: W270 – 275

Kaplowitz PB (2008) Link between body fat and the timing of puberty. *Pediatrics* 121 (Suppl 3): S208 – 217

Karges B, de Roux N (2005) Molecular genetics of isolated hypogonadotropic hypogonadism and Kallmann syndrome. *Endocr Dev* 8: 67 – 80

Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* 4: 363 – 371

Kristiansson K, Naukkarinen J, Peltonen L (2008) Isolated populations and complex disease gene identification. *Genome Biol* 9: 109

Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4: 1073 – 1081

Lin L, Conway GS, Hill NR, Dattani MT, Hindmarsh PC, Achermann JC (2006) A homozygous R262Q mutation in the gonadotropin-releasing hormone receptor presenting as constitutional delay of growth and puberty with subsequent borderline oligospermia. *J Clin Endocrinol Metab* 91: 5117 – 5121

Maggi R, Pimpinelli F, Molteni L, Milani M, Martini L, Piva F (2000) Immortalized luteinizing hormone-releasing hormone neurons show a different migratory activity in vitro. *Endocrinology* 141: 2105 – 2112

Memi F, Abe P, Cariboni A, MacKay F, Parnavelas JG, Stumm R (2013) CXC chemokine receptor 7 (CXCR7) affects the migration of GnRH neurons by regulating CXCL12 availability. *J Neurosci* 33: 17527 – 17537

Morris DH, Jones ME, Schoemaker MJ, Ashworth A, Swerdlow AJ (2011) Familial concordance for age at menarche: analyses from the Breakthrough Generations Study. *Paediatr Perinat Epidemiol* 25: 306 – 311

Mouritsen A, Aksglaede L, Sorensen K, Mogensen SS, Leffers H, Main KM, Frederiksen H, Andersson AM, Skakkebaek NE, Juul A (2010) Hypothesis: exposure to endocrine-disrupting chemicals may interfere with timing of puberty. *Int J Androl* 33: 346 – 359

Ojeda SR, Lomniczi A, Mastronardi C, Heger S, Roth C, Parent AS, Matagne V, Mungenast AE (2006) Minireview: the neuroendocrine regulation of puberty: is the time ripe for a systems biology approach? *Endocrinology* 147: 1166 – 1174

Ong KK, Elks CE, Li S, Zhao JH, Luan J, Andersen LB, Bingham SA, Brage S, Smith GD, Ekelund U *et al* (2009) Genetic variation in LIN28B is associated with the timing of puberty. *Nat Genet* 41: 729 – 733

Palevitch O, Abraham E, Borodovsky N, Levkowitz G, Zohar Y, Gothilf Y (2010) Cxcl12a-Cxcr4b signaling is important for proper development of the forebrain GnRH system in zebrafish. *Gen Comp Endocrinol* 165: 262 – 268

Palmert MR, Dunkel L (2012) Clinical practice. Delayed puberty. *N Engl J Med* 366: 443 – 453

Parent AS, Teilmann G, Juul A, Skakkebaek NE, Toppari J, Bourguignon JP (2003) The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. *Endocr Rev* 24: 668 – 693

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8: 785 – 786

Pimpinelli F, Redaelli E, Restano-Cassulini R, Curia G, Giacobini P, Cariboni A, Wanke E, Bondiolotti GP, Piva F, Maggi R (2003) Depolarization differentially affects the secretory and migratory properties of two cell lines of immortalized luteinizing hormone-releasing hormone (LHRH) neurons. *Eur J Neurosci* 18: 1410 – 1418

Pitteloud N, Meysing A, Quinton R, Acierno JS Jr, Dwyer AA, Plummer L, Fliers E, Boepple P, Hayes F, Seminara S *et al* (2006) Mutations in fibroblast growth factor receptor 1 cause Kallmann syndrome with a wide spectrum of reproductive phenotypes. *Mol Cell Endocrinol* 254–255: 60 – 69

Pitteloud N, Quinton R, Pearce S, Raivio T, Acierno J, Dwyer A, Plummer L, Hughes V, Seminara S, Cheng YZ *et al* (2007) Digenic mutations account for variable phenotypes in idiopathic hypogonadotropic hypogonadism. *J Clin Invest* 117: 457 – 463

Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A (2010) Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* 20: 110 – 121

Radovick S, Wray S, Lee E, Nicols DK, Nakayama Y, Weintraub BD, Westphal H, Cutler GB Jr, Wondisford FE (1991) Migratory arrest of gonadotropin-releasing hormone neurons in transgenic mice. *Proc Natl Acad Sci USA* 88: 3402 – 3406

Ritte R, Lukanova A, Tjonneland A, Olsen A, Overvad K, Mesrine S, Fagherazzi G, Dossus L, Teucher B, Steindorf K *et al* (2012) Height, age at menarche and risk of hormone receptor positive and negative breast cancer: a cohort study. *Int J Cancer* 132: 2619 – 2629.

Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5: 725 – 738

Saari A, Harju S, Makitie O, Saha MT, Dunkel L, Sankilampi U (2015) Systematic growth monitoring for the early detection of celiac disease in children. *JAMA Pediatr* 169: e1525

Schwarz JM, Cooper DN, Schuelke M, Seelow D (2014) MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* 11: 361 – 362

Sedlmeyer IL (2002a) Delayed puberty: analysis of a large case series from an academic center. *J Clin Endocrinol Metab* 87: 1613 – 1620

Sedlmeyer IL (2002b) Pedigree analysis of constitutional delay of growth and maturation: determination of familial aggregation and inheritance patterns. *J Clin Endocrinol Metab* 87: 5581 – 5586

Segev O, Samach A, Faerman A, Kalinski H, Beiman M, Gelfand A, Turam H, Boguslavsky S, Moshayov A, Gottlieb H *et al* (2004) CMF608-a novel mechanical strain-induced bone-specific protein expressed in early osteochondroprogenitor cells. *Bone* 34: 246 – 260

Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, Day IN, Gaunt TR (2013) Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat* 34: 57 – 65

Sun Y, Bak B, Schoenmakers N, van Trotsenburg AS, Oostdijk W, Voshol P, Cambridge E, White JK, le Tissier P, Gharavy SN *et al* (2012) Loss-of-function mutations in IGSF1 cause an X-linked syndrome of central hypothyroidism and testicular enlargement. *Nat Genet* 44: 1375 – 1381

Toba Y, Tiong JD, Ma Q, Wray S (2008) CXCR4/SDF-1 system modulates development of GnRH-1 neurons and the olfactory system. *Dev Neurobiol* 68: 487 – 503

Tobet SA, Schwarting GA (2006) Minireview: recent progress in gonadotropin-releasing hormone neuronal migration. *Endocrinology* 147: 1159 – 1165

Tomberg J, Sykiotis GP, Keefe K, Plummer L, Hoang X, Hall JE, Quinton R, Seminara SB, Hughes V, Van Vliet G *et al* (2011) Heparan sulfate 6-O-sulfotransferase 1, a gene involved in extracellular sugar modifications, is mutated in patients with idiopathic hypogonadotrophic hypogonadism. *Proc Natl Acad Sci USA* 108: 11524 – 11529

Vaaralahti K, Wehkalampi K, Tommiska J, Laitinen EM, Dunkel L, Raivio T (2011) The role of gene defects underlying isolated hypogonadotropic hypogonadism in patients with constitutional delay of growth and puberty. *Fertil Steril* 95: 2756 – 2758

Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38: e164

Wehkalampi K, Widen E, Laine T, Palotie A, Dunkel L (2008a) Association of the timing of puberty with a chromosome 2 locus. *J Clin Endocrinol Metab* 93: 4833 – 4839

Wehkalampi K, Widen E, Laine T, Palotie A, Dunkel L (2008b) Patterns of inheritance of constitutional delay of growth and puberty in families of adolescent girls and boys referred to specialist pediatric care. *J Clin Endocrinol Metab* 93: 723 – 728

Wetsel WC (1995) Immortalized hypothalamic luteinizing hormone-releasing hormone (LHRH) neurons: a new tool for dissecting the molecular and cellular basis of LHRH physiology. *Cell Mol Neurobiol* 15: 43 – 78

Widen E, Silventoinen K, Sovio U, Ripatti S, Cousminer DL, Hartikainen AL, Laitinen J, Pouta A, Kaprio J, Jarvelin MR *et al* (2012) Pubertal timing and growth influences cardiometabolic risk factors in adult males and females. *Diabetes Care* 35: 850 – 856

Wray S (2010) From nose to brain: development of gonadotrophin-releasing hormone-1 neurones. *J Neuroendocrinol* 22: 743 – 753

Zhu J, Choa RE, Guo MH, Plummer L, Buck C, Palmert MR, Hirschhorn JN, Seminara SB, Chan YM (2015) A shared genetic basis for self-limited delayed puberty and idiopathic hypogonadotropic hypogonadism. *J Clin Endocrinol Metab* 100: E646 – E654.